WIRTSCHAFTSUNIVERSITÄT WIEN
Vienna University of Economics and Business

Bachelor Thesis

English title of the Bachelor Thesis	Bayesian Model Averaging: Overview of the application in empirical growth
German title of the Bachelor Thesis	Bayesian Model Averaging: Anwendung bei empirischen Wachstumsmodellen im Überblick
Author **last name, first name(s)**	Wesely, Oliver
Student ID number	1325558
Degree program	Business, Economics and Social Science
Examiner **degree, first name(s), last name**	MMag Dr., Paul, Hofmarcher

I hereby declare that

1. I have written this Bachelor thesis independently and without the aid of unfair or unauthorized resources. Whenever content was taken directly or indirectly from other sources, this has been indicated and the source referenced.

2. this Bachelor thesis has neither previously been presented for assessment, nor has it been published.

3. this Bachelor thesis is identical with the assessed thesis and the thesis which has been submitted in electronic form.

4. (only applicable if the thesis was written by more than one author): this Bachelor thesis was written together with first name(s), last name(s). The individual contributions of each writer as well as the co-written passages have been indicated.

Date 24.04.2017

O. Wesely

Signature

Vienna University of Economics and Business

Bachelor Thesis

Bayesian Model Averaging: Overview of the application in empirical growth

Oliver Wesely
(1325558)

Degree program Business, Economics and Social Science

supervised by
MMag Dr. Paul Hofmarcher

April 24, 2017

Contents

Abstract

Since a few decades Bayesian statistics has become increasingly popular in social and behavioral science research. In this paper, I give a brief overview of current literature on Bayesian Model Averaging in the field of empirical growth, which is without any doubt the most active field in which Bayesian Model Averaging approaches are applied.

1 Introduction

Statistical methods based on the frequentist (classical) paradigm often involve testing the same null hypothesis again and again, ignoring the knowledge of previous studies. But as replication is an important and indispensable tool, Bayesian approaches integrate background knowledge into the statistical model. Therefore, the plausibility of previous findings can be evaluated in relation to new data.

The frequentist paradigm of statistics associates probability with long-run frequency and its canonical example is the notion of an infinite coin toss. A sample space of possible outcomes is enumerated, the two possible outcomes in this example are heads and tails, and probability is the proportion of the outcome over the number of coin tosses.

The Bayesian paradigm contrarily interprets probability as the subjective experience of uncertainty. Bayes' theorem is a model for learning from data, which I will show in the next section. The classical example of the subjective experience of uncertainty is the notion of placing a bet. Placing a bet, for example on a football game, involves using as much prior information as possible. Once there is new data the prior information is going to be updated.

The main difference between these two paradigms concerns the nature of unknown parameters. In frequentist statistical methods the parameter of interest is assumed to be unknown, but fixed, therefore it is assumed that there is only one true population parameter. In Bayesian statistical inference, all unknown parameters are treated as uncertain and should be described by a probability distribution.

In this paper, I particularly want to take a look at model selection in general and Bayesian Model Averaging in the field of empirical growth.

"Since it is often not clear a priori which theory is correct and which variables should be included in the 'true' regression model, a naive approach that ignores specification and data uncertainty generally results in biased parameter estimates, overconfident (too narrow) standard errors and misleading inference and predictions." [Doppelhofer and Weeks, 2009]

This problem exists in many studies and is not the uncertainty associated with the estimate conditional on a given model, which is assessed in virtually every empirical model, but it is that the empirical specification typically is taken as known. While some variations of a baseline model are often reported, standard empirical practice does not systematically account for the sensitivity of claims about the estimate of interest to model selection.

When for the inclusion q variables are available, the common situation refers to the uncertainty surrounding model selection among 2^q possible models. Imagine there are many different candidate models for estimating the effect of X on Y. Therefore, one can estimate all the candidate models and then compute a weighted average of all estimates for the coefficients of X, which is a so-called model averaging approach and considers not only the uncertainty associated with the parameter estimate conditional on a given model, but also the uncertainty of the parameter estimate across different models.

The Bayesian Model Averaging approach is an example of a model averaging approach and will be described later on. Since Bayesian statistic becomes more popular I will give a brief overview of current literature on Bayesian Model Averaging in the field of empirical growth in this paper.

In practice, model averaging methods can be implemented by using freely available software like R, where Raftery et al. and Feldkircher and Zeugner provided free R packages, BMA and BMS, for performing BMA under different priors, which I will introduce in section 7.

This paper is structured in the following way: In section 2 basic Bayesian inference theory is presented, section 3 will head on with some insights into the field of model selection, section 4 introduces the basics about Bayesian Model Averaging and in section 5 I am going to give an overview of actual literature in

Bayesian Model Averaging with applications to economics, especially to empirical growth. In section 6.1 I will introduce two R packages with Bayesian Model Averaging approaches and in the final section 7 some concluding remarks can be found.

2 Bayesian Inference

2.1 Bayes' Theorem

Theorem 1. *(Bayes' Theorem) For any two events A and B with $0 < \mathbb{P}(A) < 1$ and $\mathbb{P}(B) > 0$, we have*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

This simple formula forms the basics for the whole analysis.

2.2 Posterior Distribution

Definition 1. *(Posterior distribution) Let $X = x$ denote the observed realisation of a (possibly multivariate) random variable X with density function $f(x|\theta)$. Specifying a prior distribution with density function $f(\theta)$ allows us to compute the density function $f(\theta|x)$ of the posterior distribution using Bayes' theorem*

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}. \tag{2}$$

For discrete parameters θ, the integral in the denominator has to be replaced with a sum.

Definition 2. *(Likelihood function) The likelihood function $L(\theta)$ is the probability mass or density function of the observed data x, viewed as a function of the unknown parameter θ.*

In Bayesian inference the posterior distribution is the most important quantity, because it contains all the available information of the unknown parameter θ after having observed the data $X = x$. In (2) the term $f(x|\theta)$ is simply the likelihood function $L(\theta)$. The denominator in (2) can be written as

$$\int f(x|\theta)f(\theta)d\theta = \int f(x,\theta)d\theta = f(x), \tag{3}$$

and therefore does not depend on θ. The quantity $f(x)$ is known as the marginal likelihood.
Hence the density of the posterior distribution is proportional to the product of the likelihood and the density of the prior distribution, with proportionality constant $1/f(x)$, usually denoted as

$$\begin{aligned} & f(\theta|x) \propto f(x|\theta)f(\theta) && (4)\\ \Leftrightarrow\quad & f(\theta|x) \propto L(\theta)f(\theta) && (5)\\ \Leftrightarrow\quad & \textit{post distribution} \propto \textit{likelihood} \times \textit{prior distribution} && (6)\end{aligned}$$

where "$\propto$" stands for "is proportional to" and implies that $1/\int L(\theta)f(\theta)d\theta$ is the normalizing constant such that $\int f(\theta|x)d\theta = 1$ and therefore $f(\theta|x)$ is a valid density function.
To sum up, the density function of the posterior distribution can be retrieved from a multiplication of the likelihood function and the density function of the prior distribution with subsequent normalization.
[Held and Bové, 2014]

3 Model Selection

One of the main questions in statistics is how one selects a model from a set of candidate models for a given dataset. The central quantity for model selection is the maximal value of the likelihood function under the respective model: $max_{M_i} L(\theta)$. A more complex model will always increase the likelihood at the corresponding MLE, if the simpler model is nested, so if I would choose the model with the largest maximal likelihood I would always select the most complex one. Therefore I always have to include the dimension of models and somehow need to penalize model complexity. A more fundamental question would be if a "true" model even exists. Do I really need to select a single model at all? In the following sections I do not choose one single model, but combine results from different models with Bayesian methods, especially Bayesian Model Averaging. [Held and Bové, 2014]

3.1 Historical perspective on model combination

First of all, I would like to take a look at some historical perspectives on model combination. Almost 200 years ago Laplace already considered combining regression coefficient estimates. Therefore he derived and compared the properties of two estimators, the least squares and the weighted median.[Laplace, 1818]
Beside Laplace, a more statistical treatment of combining multiple estimates came from Edgerton and Kolbe. They combined different estimates in such a way that the combining weights result from minimizing the sum of squares of the differences of the scores. [Edgerton and Kolbe, 1936] Others derived a formula for combining multiple measures in which the criterion is obtaining maximum separation among the individual population members [Horst, 1938] or provided for example a minimum-squared-error combination of estimates [Halperin, 1961]. After the late 1970s, the idea of combining estimates was present in many studies in the field of statistics.
One of the earliest Bayesian approaches to combine estimates was presented by Geisser, Roberts and Geisel.[Geisser, 1965, Roberts, 1965, Geisel, 1973] The first comprehensive description of the basic paradigm for Bayesian Model Averaging and, therefore, the standard cited seminal paper in the Bayesian Model Averaging literature was presented by Leamer. [Leamer, 1978]

4 Bayesian Model Averaging

4.1 Estimation and Inference with BMA

In order to illustrate the model, I will consider the case of a normal linear regression model in which uncertainty comes from the selection of regressors:

$$y = X\beta + \epsilon \tag{7}$$

$$\epsilon \;\sim\; N(0, \sigma^2 I_T)$$

with $y = (y_1, \ldots, y_T)'$, ϵ are $T \times 1$ vectors, X is a $T \times q$ matrix of regressors and β contain the parameters to be estimated. If we set some components of $\beta = (\beta_1, \beta_2, \ldots, \beta_q)'$ to be zeros, we get a total of 2^q candidate models, indexed by M_j for $j = 1, \ldots, 2^q$. Each model M_j depends upon parameters β^j, following the posterior for the parameters calculated using M_j is:

$$g(\beta^j|y, M_j) = \frac{f(y|\beta^j, M_j)g(\beta^j|M_j)}{f(y|M_j)} \tag{8}$$

with a posterior $g(\beta^j|y, M_j)$, a likelihood $f(y|\beta^j, M_j)$ and a prior $g(\beta^j|M_j)$ for each model. If the prior model probability $\mathbb{P}(M_j)$ is given we can calculate the posterior model probability by using Bayes Rule:

$$\mathbb{P}(M_j|y) = \frac{f(y|M_j)\mathbb{P}(M_j)}{f(y)} \tag{9}$$

According to 8 and 9 it is clear that we now need to elicit priors for the parameters of each model and for the model probability itself. Following Leamer we can consider β as a function of β^j and calculate, under consideration by the law of total probability, the posterior density of the parameters for all the models:

$$g(\beta|y) = \sum_{j=1}^{2^q} \mathbb{P}(M_j|y) g(\beta|y, M_j) \tag{10}$$

Hence the full posterior distribution of β is a weighted average of its posterior distributions with $\mathbb{P}(M_j|y)$ given weights. This is framed as Bayesian Model Average (BMA). When applying the equation 10, estimation and inference come together naturally from the posterior distribution, which provides inference about β that takes full account of model uncertainty. We also can easily compute the model-averaged posterior expectation of β and their associated variances:

$$\begin{aligned}
\mathbb{E}(\beta|y) \quad &= \sum_{j=1}^{2^q} \mathbb{P}(M_j|y)\mathbb{E}(\beta|y, M_j) \\
\mathbb{V}(\beta|y) \quad &= \sum_{j=1}^{2^q} \mathbb{P}(M_j|y)\mathbb{V}(\beta|y, M_j) + \sum_{j=1}^{2^q} \mathbb{P}(M_j|y)(\mathbb{E}(\beta|y, M_j) - \mathbb{E}(\beta|y))^2
\end{aligned}$$

Implementing BMA can be difficult because of two reasons:

1. Priors on parameters and on models need to be elicited for many models, which can be a complicated task,
2. 2^q is the number of models under consideration and, therefore, it is often so large that the computational burden of BMA might be prohibitive.

In order to find solutions to these challenges, I present some of the remedies proposed in the literature. [Moral-Benito, 2015]

4.2 Priors

4.2.1 Priors on the Parameter Space

For Bayesian model averaging we need prior distributions for the unknown parameters under the different models M_j for $j = 1, \ldots, 2^q$, even when we only have few prior knowledge to elicit them. If a researcher has only little information about the unknown parameters, improper priors have been commonly employed as representations of this ignorance, which are priors that do not integrate to one. As there might be some difficulties in working with non-informative improper priors, mainstream priors in BMA include hierarchical prior structures involving improper priors for the common parameters, and proper priors for the remaining parameters. [Moral-Benito, 2015]
BMA literature commonly uses the following priors:

1. Improper Priors
2. Zellner's g Priors
3. Laplace Priors

4. Empirical Bayes (EB) Priors

The usage of improper priors is getting a problem when comparing different models, as we do with the BMA approach. To prevent this difficulty in using improper priors, various approaches have been advocated, for example the "imaginary training sample device" of Spiegelhalter and Smith and one other from Lempers. [Spiegelhalter and Smith, 1982, Lempers, 1971]
Given the difficulty of implementing these methods to different competing models simultaneously, BMA is often based on proper and sometimes on non-informative priors.
One popular prior is the Zellner's g Prior. Given a Gaussian regression model Zellner suggested a particular form of the conjugate Normal-Gamma family, namely a g prior:

$$p(\phi) \propto \frac{1}{\phi}, \qquad \beta|\phi \sim N(\beta_a, \frac{g}{\phi}(X^T X)^{-1}),$$

where the prior mean β_a is taken as the anticipated value of β based on imaginary data and the prior covariance matrix of β is a scalar multiple g of the Fisher information matrix, which depends on the observed data through the regressor matrix X. In the context of hypothesis testing with $H_0 : \beta = \beta_0$ versus $H_1 : \beta \in \mathbb{R}^k$, Zellner suggested setting $\beta_a = \beta_0$ in the g prior for β under H_1 and derived expressions for the Bayes factor for testing H_1 versus H_0.
The simplicity of the g-prior formulation is that just the parameter g needs to be specified. Because g acts as a dimensionality penalty the choice of g is critical. [Liang et al., 2008]
Under uniform prior model probabilities, the choice of g effectively controls model selection, with large g typically concentrating the prior on parsimonious models with a few large coefficients, whereas small g tends to concentrate the prior on saturated models with small coefficients. [George and Foster, 2000]
Three of the most popular alternatives for g are:

- *Unit Information Prior.* Kass and Wasserman recommended choosing priors with the amount of information about the parameter equal to the amount of information contained in one observation. Therefore, the unit information prior corresponds to taking $g = n$, leading to Bayes factors that behave like the BIC. [Kass and Wasserman, 1995]
- *Risk inflation criterion.* Foster and George recommended the use of $g = p^2$ from a minimax perspective. [Foster and George, 1994]
- *Benchmark prior.* Fernández et al. determined the combination of "Unit Information Prior" and "Risk inflation criterion" priors to perform best with respect to predictive performance with the recommendation to take $g = max(n, p^2)$. [Fernández et al., 2001a]

In order to illustrate Laplace priors let us construct a partition of the X matrix, such that we can rewrite 7 as follows:

$$y = X_1\gamma + X_2\delta + \epsilon \tag{11}$$

$$\epsilon \ \sim \ N(0, \sigma^2 I_N)$$

with γ ($q_1 \times 1$) and δ ($q_2 \times 1$) parameter vectors with $q_1 + q_2 = q$. Model 11 can be reparametrized replacing $X_2\delta = X_2^*\delta^*$ with $X_2^* = X_2 P \pi^{-1/2}$ and $\delta^* = \pi^{1/2} P'\delta$, where P is an orthogonal matrix and π is a diagonal matrix such that $P'X_2'R_{X_1}X_2P$ and $R_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$. With theses settings, an alternative prior structure has been considered by Magnus et al. which leads to the so-called weighted-average least squares (WALS) estimator. On the one hand, the WALS estimator employs non-informative model-specific priors and remarkably reduces the computational burden of standard BMA being proportional to q_2 (or q) instead of 2^{q_2} (or 2^q). Yet, on the other hand, WALS does not provide neither Bayesian posterior distributions nor posterior inclusion probabilities as a measure of robustness. [Moral-Benito, 2015]

And last, I will take a look at the empirical Bayes (EB) priors, which in contrast estimate the hyperparameter g from the data, rather than pre-select g a priori. In the literature two common EB alternatives are convenient, the local EB approach by Hansen and Yu [Hansen and Yu, 2001] and the global EB approach by George and Foster [George and Foster, 2000], and Clyde and George [Clyde and George, 2000]. But it is a fact that Bayesians in general are critical with this approach, as it does not correspond to a formal Bayesian procedure. [Moral-Benito, 2015]

4.2.2 Priors on the Model Space

For implementing any of the described BMA strategies, prior model probabilities ($\mathbb{P}(M_j)$) must be assigned. In BMA research, priors like the Binomial, Binomial-Beta and Dilution priors are commonly used. The most common prior structure in BMA research is the Binomial distribution following Moral-Benito, where each variable is independently included (or not) in a model so that the model size follows a Binomial distribution with a specific probability of success [Moral-Benito, 2015]. Binomial-Beta priors in which the probability of success is treated as random were proposed by Ley and Steel. This hierarchical prior implies a substantial increase in prior uncertainty about the model size and makes the choice of prior model probabilities much less critical. [Ley and Steel, 2009]
The Binomial and the Binomial-Beta priors have the implicit assumption, that the probability of one regressor appearing in the model is independent of the inclusion of others, in common. With this priors on model space, any researcher could arbitrarily modify the prior model probabilities across theories simply by including redundant proxy variables for some of these theories, which is called the denominated dilution problem raised by George [George, 1999].
To address this issue, Durlauf et al [Durlauf et al., 2008] introduced a version of George' dilution priors that assigns probability to neighborhoods of models. Moreover, this kind of dilution prior assigns uniform probability to neighborhoods rather than to models, and solves the dilution problem. [Moral-Benito, 2015]
In which other situations will dilution priors be useful? Dilution priors avoid placing too little probability on good and unique models as a consequence of massing excess probability on large sets of bad and similar models. Thus, dilution priors may be useful for model averaging over the entire posterior to avoid biasing averages away from good, but isolated models. They also may be useful for Markov Chain Monte Carlo (MCMC) sampling because such Markov chains gravitate toward regions of high probability. Failure to dilute the probability across clusters of many bad models would bias both model search and model averaging approximations toward those bad models. Another thing which should be noted is that dilution priors would not be appropriate for pairwise model comparisons because the relative strengths of two models should not depend on whether another is considered. For this purpose, George would preferably use Bayes factors. [George, 2010]
Although dilution priors may occasionally arise, the construction of dilution priors seems to require at least a bit more crafting, therefore, I would like to refer to sections $3-5$ in George [George, 2010].

4.3 Further Topics in BMA

4.3.1 Computational aspects

Although we theoretically should be able to execute the BMA with the results described above, practically the number of models under consideration (2^q) can be really big, hence it is often impossible to estimate every single model. Regarding this computational problem, several algorithms have been developed in the literature, which carry out BMA without evaluating every possible model. [Moral-Benito, 2015]
The two commonly used approaches are the so-called Occam's Window proposed by Madigan and Raftery [Madigan and Raftery, 1994], which exclude models from the summation, that predict the data far less well than the best model and models that receive less support than any of their simpler submodels, and MC which was initially develpoed by Madigan and York [Madigan and York, 1995]. MCMC methods take draws from the parameter space in order to simulate the posterior distribution of interest and focus on regions of high posterior probability. BMA considers the models as discrete random variables so that posterior simulators, which draw from the model space instead of the parameter space, can be derived.

4.3.2 Jointness

Another relevant topic in the BMA framework is the jointness among explanatory variables, which tells us whether different sets of regressors are substitutes or complements in the determination of the outcome and in the context of BMA. Therefore Ley and Steel [Ley and Steel, 2007] and Doppelhofer and Weeks [Doppelhofer and Weeks, 2009] define ex post jointness measures that appear in linear regression models. Both approaches are interested in the measure of two regressors X_i and X_j in the context of linear regressions. In this section, I am going to take a look at the related work of Ley and Steel and only want to refer to the paper of Doppelhofer and Weeks.
Ley and Steel proposed four criteria that a useful measure of jointness should have:

- *Interpretability.* Any jointness measure should have either a formal statistical or a clear intuitive meaning in terms of jointness.
- *Calibration.* Values of the jointness measure should be calibrated against some clearly defined scale, derived from either formal statistical or intuitive arguments.
- *Extreme jointness.* The situation where two variables always appear together should lead to the jointness measure reaching its value reflecting maximum jointness.
- *Definition.* The jointness measure should always be defined whenever at least one of the variables considered is included with positive probability.

Consider the variables i and j and let $\mathbb{P}(i)$ denote the posterior probability of inclusion for regressor i, which means $\mathbb{P}(i)$ is defined as the sum of posterior probabilities of all models that contain regressor i. Following $\mathbb{P}(i \cap j)$ is the posterior probability of including both variables i and j following $\mathbb{P}(i) \geq \mathbb{P}(i \cap j)$. A raw measure of jointness would simply be the probability of joint inclusion, $\mathbb{P}(i \cap j)$. But two better alternative measures of jointness are proposed in their paper and surely satisfy all the proposed criteria. The first one is the joint probability relative to the probability of including either one:

$$\mathcal{J}^*_{ij} = \frac{\mathbb{P}(i \cap j)}{\mathbb{P}(i \cup j)} = \frac{\mathbb{P}(i \cap j)}{\mathbb{P}(i) + \mathbb{P}(j) - \mathbb{P}(i \cap j)} \in [0, 1] \tag{12}$$

and the other one the joint probability relative to the probability of including either one, but not both, excluding the intersection itself:

$$\mathcal{J}_{ij} = \frac{\mathbb{P}(i \cap j)}{\mathbb{P}(i \cap \tilde{j}) + \mathbb{P}(j \cap \tilde{i})} = \frac{\mathbb{P}(i \cap j)}{\mathbb{P}(i) + \mathbb{P}(j) - 2\mathbb{P}(i \cap j)} \in [0, \infty) \tag{13}$$

where $\tilde{i}$ and $\tilde{j}$ stand for the exclusion of i and j, respectively. Both measures of jointness surely satisfy the criteria proposed in this paper.
Another big advantage of these measures is that they can be easily extended in the case of more than two regressors, the so-called multivariate jointness. This can be defined for general sets of regressors $\mathcal{S}$ through two quantities: $\mathbb{P}(\mathcal{S})$, the total posterior probability assigned to those models having all regressors in $\mathcal{S}$, and $\mathbb{P}(\subset \mathcal{S})$, defined as the posterior mass assigned to all models including only proper subsets of $\mathcal{S}$. With these definitions we can generalize the measures above with:

$$\mathcal{J}^*_{\mathcal{S}} = \frac{\mathbb{P}(\mathcal{S})}{\mathbb{P}(\subset \mathcal{S}) + \mathbb{P}(\mathcal{S})} \in [0, 1] \tag{14}$$

$$\mathcal{J}_{\mathcal{S}} = \frac{\mathbb{P}(\mathcal{S})}{\mathbb{P}(\subset \mathcal{S})} \in [0, \infty). \tag{15}$$

[Ley and Steel, 2007]

5 Model averaging with applications to economics

5.1 A brief overview of the literature

Until the late 1990s and 2000s, the BMA approach was mostly ignored in economic applications until the 'BMA revolution' in economics took place. Only Geisel who compared the prediction ability of macro-models based on posterior model probabilities of consumption equations [Geisel, 1973], and Moulton who applied model averaging to 4096 hedonic regressions of quality-adjusted price index numbers of radio services [Moulton, 1991], represent the two exceptions of economic research considering model averaging previous to the 'BMA revolution'.

One of the first who included the model uncertainty and tried to highlight the influence of rethinking how to formulate and present policy advice in economics was Brock [Brock et al., 2003, Brock et al., 2006].

Model averaging has been considered in many fields of economics such as forecasting output growth by Koop and Potter [Koop and Potter, 2004], exchange rates by Wright [Wright, 2008a], predicting the inflation and output growth in the UK by Garratt et al. [Garratt et al., 2009] and forecasting the US inflation rate by Wright [Wright, 2008b]. Regarding labour and migration, Koske and Wanner [Koske and Wanner, 2013] studied the drivers of income inequality in OECD countries and Mitchell et al. [Mitchell et al., 2011] investigated the determinants of international migration to the UK combining panel data with model averaging techniques. In the field of finance Pesaran et al [Pesaran et al., 2009] for example employed model averaging as a remedy to the risk of inadvertently using false models in portfolio management. Based on model averaging Avramov and Cremers [Avramov, 2002, Cremers, 2002] predicted stock returns and reported improved forecasting performance of the BMA approach. In other fields of economics for example Crespo-Cuaresman and Slacik [Crespo Cuaresma and Slacik, 2009] identified the most important determinants of currency crisis in the framework of binary choice models for a panel of countries, in health economics Morales et al. [Morales et al., 2006] characterized the dose-response relationship between an environmental exposure and adverse health outcomes using model averaging techniques. [Moral-Benito, 2015]

Nevertheless, the field of empirical growth is the one in which model averaging is applied the most. Therefore, I am going to take a closer look at this field in the following section.

5.2 Literature overview of model averaging in the field of empirical growth

Without any doubt, empirical growth is one of the most active fields in which model averaging approaches are applied.

One of the first seminal paper, which tries to examine whether the conclusions from existing studies are robust or fragile to small changes in the conditioning information set, was written by Ross Levine and David Renelt [Levine and Renelt, 1992] (henceforth LRD).

In this paper, the authors are using cross-country regressions to search for empirical linkages between long-run growth rates and other indicators of economic policy and politics and are trying to find if conclusions are robust or fragile. In most of the existing studies, investigators consider only a small number of explanatory variables in attempting to establish a statistically significant relationship between growth and a particular variable of interest. But given that more than 50 variables have been found to be significantly correlated with growth in at least one regression a big uncertainty of readers, regarding the confidence they should place in the results of any study, would arise. Therefore, the paper had the aim to address the question: *How much confidence should we have in the conclusions of cross-country growth regressions?*.

To test the robustness of coefficient estimates to alterations in the conditioning set of information they use a variant of Edward E. Leamer's extreme-bounds analysis [Leamer, 1983]. They studied the statistical relationship between growth and a wide array of variables and indicators in their paper. The relationship of a particular variable and growth is considered robust if it remains statistically significant. Whereas, the relationship of the theoretically predicted sign is considered as robust if the conditioning set of variables in the regression changes.

The results showed that almost all identified relationships are sensitive to alterations in the conditioning set of variables and change sign with small changes in the conditioning set of variables. Their analysis

also found some robust relationships, for example between average growth rates and the average share of investment in GDP. They also found that the ratio of trade to output is robustly, positively correlated with the investment share. Their main finding is that the cross-country statistical relationship between long-run average growth rates and almost every particular macroeconomic or any other economic, political indicator is fragile. [Levine and Renelt, 1992]
With Leamer's extreme-bounds test to identify "robust" empirical relations the problem is, that if one finds a single regression for which the sign of the coefficient changes or becomes insignificant the variable is not robust. Which not surprisingly follows, that LRDs' conclusion is, that most or even all of the variables are not robust. One possible reason for this might be that very few variables can be identified to be correlated with growth, hence, that some researchers concluded that nothing can be learned from LRDs' paper because no variables are robustly correlated with growth. Another reason might be that the test is too strong for any variable to pass it.
Finding another test, which is not a so called "extreme test" labeling variables as "robust" vs. "non-robust", was tried by Sala-I-Martin [Sala-i Martin, 1992] (henceforth SIM). Instead of the "extreme test", he wants to assign some level of confidence to each of the variables. Therefore, he looked at the entire distribution of the estimators of a variable, but the immediate problem was that, even though each individual estimate follows a Student-t distribution, the estimates themselves could be scattered around in a strange fashion. Hence that he operates under the two different assumptions that the distribution of the estimates of a variable across models is normal or not normal.
As a result, a substantial number of variables can be found to be strongly related to growth. These variables which appear to have a "significant" effect include:

- *Regional variables:* Sub-Saharan Africa and Latin America are negatively related and Absolute Latitude tells us, that if a country is far away from the equator it is good for growth.
- *Political variables:* Rule of Law, Political Rights and Civil Liberties are good for growth and number of revolutions, and military coups and a War dummy is bad for growth.
- *Religious variables:* Confucian, Buddhist and Muslim are positive and Protestant and Catholic are negative.
- *Market Distortions and Market Performance:* Real Exchange Rate Distortions and Standard Deviation of the Black Market Premium are both negative.
- *Types of Investment:* Equipment and Non-Equipment Investment are both positive.
- *Openness:* Number of years an economy has been open between 1950 and 1990 is positive.
- *Type of Economic Organization.* Degree of capitalism is positve.

[Sala-i Martin, 1992]
In response to LRD and SIM Fernández et al. [Fernández et al., 2001b] (henceforth FLS) investigated the issue of model uncertainty in cross-country growth regressions using BMA. In particular, a Bayesian framework allowed them to deal with both, model and parameter uncertainty, in a straightforward and formal way and considered a large set of possible models by allowing for any subset of up to 41 regressors to be included. To solve this numerical problem they used an adaption of the MCMC techniques.
FLSs' findings, based on the same data as SIM, broadly support the more "optimistic" conclusions of SIM than those of LRD. They namely would conclude that some variables are important regressors for explaining cross-country growth patterns. But the identified variables are different to those of SIM. Another important difference is that in this study they do not advocate selecting a subset of the regressors and instead use BMA as described in section 4, hence, that their methodology allows them to go substantially further than previous studies.
The results of FLS agree with SIM that some regressors can be identified as useful explanatory variables for growth in a linear regression model, but they advocate a formal treatment of model uncertainty. From

the huge spread of the posterior mass in model space and the predictive advantage of BMA, it is clear that model averaging is recommended when dealing with growth regressions. Even though they find a roughly similar set of variables that might be interpreted as "important" for growth regressions, a crucial additional advantage is that their results are immediately interpretable in terms of model probabilities and all inference can easily be conducted in a purely formal fashion by BMA. In their point of view, the treatment of a very large model set in a theoretically sound and empirically practical fashion requires BMA and MCMC methods. Additionally, this methodology provides a clear and precise interpretation of the results and immediately leads to posterior and predictive inference. [Fernández et al., 2001b]

Regarding the robustness of explanatory variables in cross-country economic growth regressions, a novel approach was introduced and employed by Sala-I-Martin et al. [Sala-i Martin et al., 2004] (henceforth SEA). This approach constructs estimates by averaging ordinary least squares (OLS) coefficients across models, which they call *Bayesian Averaging of Classical Estimates* (BACE). BACE combines the averaging of estimates across models, which is a Bayesian concept, with classical OLS estimation which comes from the assumption of diffuse priors. In contrast to averaging across models, which is an inherently Bayesian idea, BACE limits the effect of prior information and uses an approach familiar to classical econometricians. Their BACE approach has several important advantages over previously used model-averaging and robustness-checking methods:

- In contrast to a standard Bayesian approach that requires the specification of a prior distribution for all parameters, BACE requires the specification of only one prior hyper-parameter, the expected model size.
- The interpretation of estimates is straightforward for economists not trained in Bayesian inference, as the weights applied to different models are proportional to the logarithm of the likelihood function corrected for degrees of freedom.
- Their estimates can be calculated using only repeated applications of OLS.
- In contrast to LRD and SIM they consider models of all sizes and no variables are held "fixed" and therefore "untested".
- They calculate the entire distribution of coefficients across models and do not focus solely on the bound of the distribution.

In the paper, they find striking and surprisingly clear conclusions, identifying a set of 18 variables as significantly related to economic growth, which have a great deal of explanatory power and are very precisely estimated. In general, their results support SIM rather than LRD. The strongest evidence is found for primary schooling enrollment, the relative price of investment goods and the initial level of income, where the latter reflects the concept of conditional convergence. Other important variables include regional dummies such as Sub-Saharan Africa or Latin America, measures of human capital and health, religious dummies and some sectoral variables like mining. The public consumption and public investment shares are negatively related to growth, although the results are significant only for certain prior model sizes. In addition, SEAs' results are quite robust to the choice of the prior model size, which means that most of the "significant" variables in the baseline case remain significant for other prior model sizes. [Sala-i Martin et al., 2004]

Performing inference on the determinants of GDP growth is also 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]. In practice, as we already have seen, 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 in the model Ley E. and Steel M.F. [Ley and Steel, 2007] (henceforth LSM7) call it *jointness*, whereas if they perform very similar roles they should not appear jointly, which they denote by *disjointness*.

Various approaches to deal with this model uncertainty have appeared in the literature and this paper (LSM7) propose alternative measures of jointness for Bayesian variable selection, based on probabilistic arguments and illustrate their application in the context of BMA, as we already explained in section 4.3.2.

LSM7 applied their jointness measure to two data sets and encountered jointness only between important determinants of growth and situations of disjointness where regressors are substitutes and really should not appear together. The regressors displaying disjointness relationships tend to be fairly unimportant drivers of growth. Considering triplets rather than pairs of variables these conclusions are strengthened and even less jointness and more disjointness can be found.

Thus, the data sets analyzed in the paper seem to contain a few key growth determinants that have a clearly defined and separate role to play, 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 they 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. [Ley and Steel, 2007]

Prior assumptions can also be extremely critical for the outcome of BMA analysis. Therefore, Ley E. and Steel M.F. [Ley and Steel, 2009] (henceforth LSM9) showed in detail how the prior assumptions affect their inference and that finding similar posterior results on inclusion probabilities of regressors, with quite different prior settings and using the same data, mostly results by accident. They focus on a general prior structure which encompasses most priors used in the growth regression literature and allows for prior choice in the two areas: the choice of the precision factor g in the g-prior and the prior assumptions on the model space. They elicit the prior in terms of the prior mean model size as other aspects of the prior are typically less interpretable for most applied analysts.

Using three different datasets that have been used in the literature, LSM9 assess the effect of prior setting for posterior inference on model size, but they also consider the spread of model probabilities over the model space and the posterior inclusion probabilities of the regressors. The key motivation for this analysis is often the relative importance of the regressors as drivers of growth. With repeatedly splitting the samples into an inference part and a prediction part they try to examine the robustness of the inference with respect to changes of the dataset.

LSM9s' concluding remarks include that prior assumptions for BMA are critically important and it clearly matters what prior settings are chosen. They would recommend using a random θ rather than a fixed one, since the hierarchical prior is much less sensitive to the choice of prior mean model size. They strongly advise against choosing $g = 1/n$ with fixed $\theta = 0.5$ as this combination may lead to relatively bad predictions. In general they recommend avoiding the choice of $g = 1/n$, which implies a fairly small model size penalty and can result in convergence problems and has also displayed more sensitivity to the mean model size than the alternative $g = 1/k^2$. [Ley and Steel, 2009]

Three years later Ley E. and Steel M.F. [Ley and Steel, 2012] (henceforth LSM12) issued a seminal paper again about the effect of the prior on the results, such as posterior inclusion probabilities of regressors and predictive performance.

LSM12 combine a Binomial-Beta prior on model size with a g-prior on the coefficients of each model. In addition, they assign a hyperprior to g, as the choice of g has been found to have a large impact on the results. For the prior on g, they examine the Zellner-Siow prior and a class of Beta shrinkage priors. They propose a benchmark Beta prior, inspired by earlier findings with fixed g, and show it leads to consistent model selection.

Previous studies of variable selection in linear regression using g-priors has been noted that the choice of g is crucial for the behaviour of BMA procedures and that the prior on the model space is an important element of the model, particularly in the way it penalizes larger models. If a priori each covariate is included independently with probability θ in the model, the interaction between θ and g was already explored in some detail in LSE9, where the use of a hierarchical prior on θ was recommended, as a way to make the analysis more robust with respect to prior assumptions on the model space.

LSE12 go one step further and the hierarchical Bayesian model explored of them has a hyperprior on θ (which leads to an integral to compute prior model probabilities, which can fortunately be solved analytically) and a hyperprior on g, which leads to an integral for the marginal likelihood that is solved by adding g into the MCMC procedure by an extra Metropolis-within-Gibbs step. Based on earlier recommendations for fixed values of g, they propose a benchmark Beta class of priors and investigate its properties. They investigate the behavior of the various priors in BMA with simulated data and various different sets of real data relating to economic applications with two sets of macroeconomic growth data and one data set regarding returns to education.

In LSE12s' concluding remarks and recommendations, the two priors that stand out by not having displayed any truly bad behaviour in their experiments are the benchmark Beta prior and the hyper-g/n prior. Thus, these priors provide an interesting compromise and would be their general recommendations to practitioners. [Ley and Steel, 2012]

After looking at the two papers on the critical prior assumption, the potential problem of multicollinearity needs to be addressed among the set of potential explanatory variables, which most of the existing methods do not assess explicitly. Yet, Hofmarcher et al. [Hofmarcher et al., 2015] (henceforth HEA) precisely look at this problem.

In their paper, they evaluate a large number of available Bayesian methods aimed at dealing with the problem of model uncertainty in the presence of correlated regressors. In addition to standard Gaussian g-priors, they also assess prior structures that have been proposed in the framework of Bayesian bridge regression, which allow them to deal explicitly with the problem of correlated explanatory variables by shrinking coefficients. Together with the most prominent cases of the bridge regression class (ridge regression and LASSO) they also include Bayesian elastic net specifications in their analysis. Furthermore, the use of a spike and slab prior allows them to include explicitly prior information concerning model size or the relative importance of covariates in the specification in a straightforward manner.

In order to assess quantitatively the importance of accounting for correlated regressors under model uncertainty in the setting of cross-country growth regressions, they provide a thorough comparison of Bayesian shrinkage methods in terms of out-of-sample predictive accuracy. They evaluate the relative predictive ability of different methods proposed in the literature making use of the dataset in SEA.

Their concluding results indicate that Bayesian LASSO, Bayesian bridge regression and Bayesian elastic nets present better out-of-sample prediction properties than standard model averaging methods, which do not explicitly account for shrinkage in individual specifications beyond the penalty implied by the posterior model probability when Zellner's g-priors are used. [Hofmarcher et al., 2015]

Previous work assessing joint covariate inclusion in BMA applications has focused on capturing relevant dependency structures using bivariate measures, that is, concentrating on the analysis of the joint posterior distribution of the inclusion of pairs of variables over the modelspace. LSM7 and LSM9 offer alternative measures of jointness. LSM7 particularly formulate a set of desirable properties for jointness measures.

An alternative approach was proposed by Crespo Cuaresma et al. [Crespo Cuaresma et al., 2016] (henceforth CEA16) using latent class analysis (LCA). Their aim was at succinctly and comprehensibly describing the dependency structure across variables in the model space using LCA (see [Vermunt and Magidson, 2002]) and applying it to economic growth regressions.

The main idea behind LCA is to relate the realizations of observed variables to an unobserved, categorical latent variable which captures the dependency structure between the observed variables. This latent variable groups observations in such a way that the dependency between variables is reduced to a minimum within groups. By applying LCA methods to the covariate inclusion structure of best models identified by BMA, we are able to capture the dependency patterns across included covariates through a (unobserved) latent variable, which induces classes with independent covariates conditional on class membership.

CEA16 proposed a method that provides a tool for applied econometricians that goes beyond the identification of individual robust determinants of socioeconomic variables by distilling the joint covariate structures that underlie the distribution of the posterior model probability across specifications. They apply their approach to the two widely known datasets of FLS and SEA for assessing the robustness of economic growth determinants.

Their results in the paper indicate that within the set of models sampled by the Markov chain in the BMA analysis of determinants of economic growth, several distinct clusters of models by covariate inclusion can be identified. For the FLS data, they identify seven clusters of models which differ in the inclusion structure for geographic, institutional and religious covariates. In contrast, the SEA dataset only reveals three latent classes with very different dependency structures. The inclusion of the variable measuring malaria prevalence

is shown to vary strongly across clusters, with its effect on economic growth being captured often by other factors such as the fraction of tropical area and coastal population density. [Crespo Cuaresma et al., 2016]
Last but not least in contrast to the previous seminal papers, Masanjala W.H. and Papageorgiou C. [Masanjala and Papageorgiou, 2008] (henceforth MP) try to find explanatory variables which can effectively explain only Africa's growth by using BMA and estimating based on a subset of the SIMs' dataset. Their baseline dataset includes 24 regressors for 104 countries (37 sub-Sahara African countries) where most of the data points represent a cross-section of average values measured over the 1960 – 1992 period and the dependent variable is represented by per capita GDP growth. Some important trends that can be found in the data are for example that Africa appears to have started from a more disadvantaged position than the rest of the world, life expectancy at birth was only 41 years compared to 59 years in the rest of the world in 1960. The primary school enrollment was only 42% compared to 84% elsewhere. Africa is also geographically disadvantaged as its larger fraction of land lies in the tropics and it has a higher degree of ethnolinguistic fractionalization. Additionally, their descriptive statistics showed that African citizens enjoyed a lower level of political right and civil liberties and that African countries were more likely to change holders of executive office through unconstitutional means (revolutions and coups).
MPs' conclusion is that relevant growth variables can be quite different for Africa. In contrast to the rest of the world, some posterior coefficient estimates reveal that key engines of growth in Africa are extremely different. Sub-Saharan growth is shown to be much more closely associated with the share of the economy made up by primary commodity production and primary education. They also show that the share of mining is a robust and positive determinant but not for the rest of the world. The differential effect of primary commodity exports, primary education and mining raise new questions about African development: Why do these variables in particular have distinct growth effects in sub-Saharan Africa? [Masanjala and Papageorgiou, 2008]
But shortly after MP, Crespo Cuaresma J. [Crespo Cuaresma, 2010] (henceforth CCJ) wrote a comment on their paper which shows that their results are not robust to the use of a prior over the model space including interaction terms. MP assume a unit information prior on the parameters of a given model and a uniform prior on the model space. In MP's contribution, no special treatment is given to the interaction terms, which are handled as if they were just other potential regressors in the set of all potential growth covariates. This might not be the most adequate choice since the interpretation of an interaction term parameter requires that the effect is exclusive to that particular product of covariates and not driven by the independent effects of the interacted variables. The real problem in MP highlighted by CCJ lies at the very heart of the interpretation of interaction effects in regression models. In a standard BMA design, such as the one in MP, where interactions are just treated as additional potential covariates only 25% of the specifications where an interaction appears can be properly interpreted, in the sense that the model also contains linear terms of the interacted variables. From a statistical point of view MP's approach is fundamentally not wrong, but when it comes to interpreting the results, some difficulties may arise concerning the evidence for parameter heterogeneity if all possible combinations of variables and interactions are allowed to specify a model.
Models including the interaction of the African dummy with a correlated variable will show an effect that may not be due to the interaction itself, but just be related to the unobservables that the dummy variable would be capturing. This problem is related to that of how to deal with correlated variables in BMA. A possible way to account for this problem in the presence of interaction terms would be to allow exclusively for either model without interactions or models where the interaction term is evaluated after controlling for both the African dummy and the independent effect of the variable being interacted.
In this comment of CCJ, they replicate MP's analysis using a prior over the model space that respects Chipman's [Chipman, 1996] strong heredity principle and shows that MP's results are not robust to this prior over the model space. Their results also show that there is no clear evidence of parameter heterogeneity between African economies and the rest of the world when this prior is used. Furthermore, they show that BMA based on the strong heredity prior leads to better out-of-sample predictive results than using the MP prior.[Crespo Cuaresma, 2010]
This example should show us that we always have to take a precise look at the used methodology of other papers and not just take the results without thinking about the underlying assumptions.

6 Introducing two R packages with Bayesian Model Averaging approaches on two datasets

6.1 Attitude data set

In this section I will use the attitude dataset, a small built in dataset of R, to show the differences in running BMA approaches between the packages BMA and BMS. The attitude dataset includes 6 variables and, therefore, $2^6 = 64$ possible model combinations.
First I will use the R package BMA [Raftery et al., 2015] to run a BMA approach. To perform BMA you have to load the BMA package first:

```
R> library("BMA")
```

and perform BMA via the function `bicreg()` and write results into the variable `att`, the first parameter of the function should be a matrix of independent variables and the second one a vector of values for the dependent variable:

```
R> att<-bicreg(attitude[,2:6],attitude[,1])
```

The bicreg function does Bayesian model selection and accounts for model uncertainty in linear regression models using the Bayesian Information criterion (BIC) approximation. For more details about the theoretical background of this package and about its functions I refer to the paper of Hoeting et al. [Hoeting et al., 1999]. After computing the model selection, the coefficient results can be obtained via

```
R> summary(att)
```

with its output in table 1.

Table 1: Attitude dataset, BMA package, summary(att)

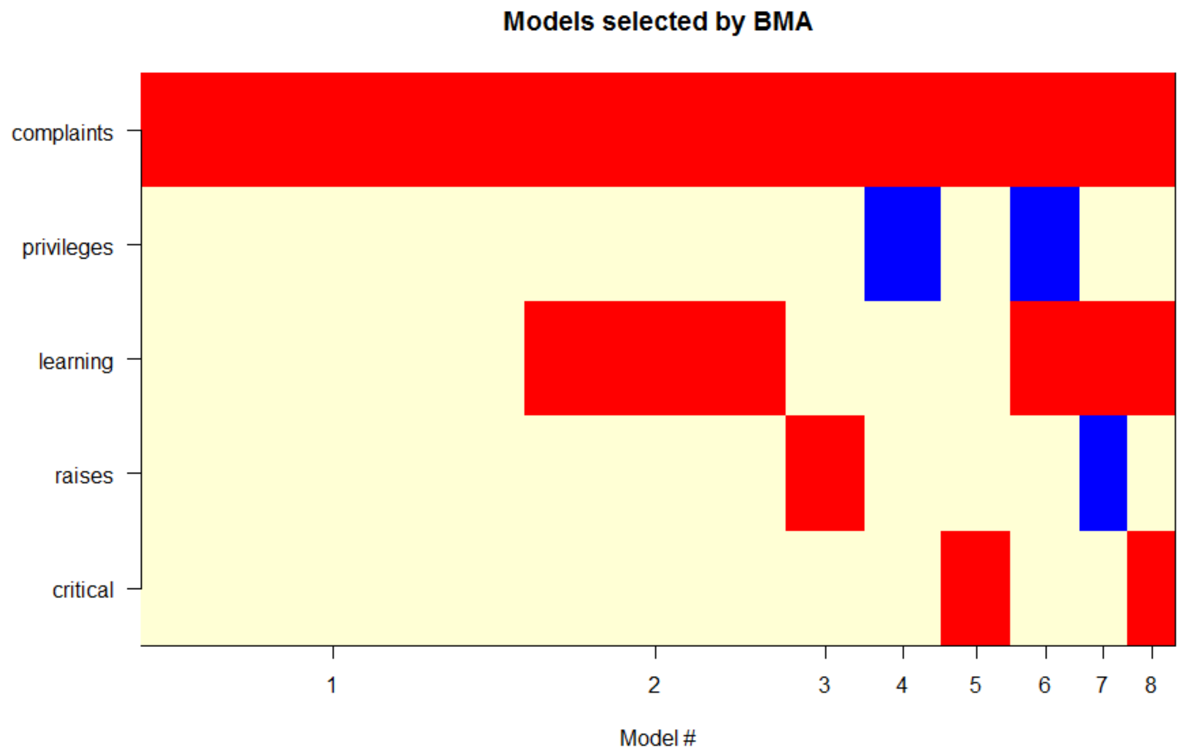
	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	12.523847	7.97426	14.376319	9.870880	11.987317	15.327620	14.251378
complaints	100.0	0.710683	0.12479	0.754610	0.643518	0.712760	0.780343	0.754344
privileges	14.0	-0.010523	0.05594	.	.	.	-0.050160	.
learning	41.1	0.088956	0.13834	.	0.211192	.	.	.
raises	12.3	0.004799	0.06546	.	.	0.080085	.	.
critical	11.4	0.000170	0.04545	.	.	.	.	0.001908
nVar				1	2	2	2	2
r2				0.681	0.708	0.684	0.683	0.681
BIC				-30.904896	-30.129704	-27.748505	-27.668889	-27.504640
post prob				0.371	0.252	0.077	0.074	0.068

The column headed ”p!=0” shows the posterior probability if the variable is included in the model (in %). The column headed ”EV” shows the BMA posterior mean, and the column headed ”SD” shows the BMA posterior standard deviation for each variable. The following five columns show the parameter estimates of the best five models found, together with the numbers of variables they include, their R^2 values, their BIC values and their posterior model probabilities.
In table 1 you can see that especially the `complaints` variable has a high posterior probability of being in the model (100%). The second best model also includes variable `learning` which has a posterior probability of being in the model of 41.1%.
A visual summary of the BMA output is produced by the `imageplot.bma` command:

```
R> imageplot.bma(att)
```

shown in figure 1.

Figure 1: BMA package, attitude dataset



Each row corresponds to a variable, and each column corresponds to a model. The corresponding rectangle is blue in the case of a positive coefficient, red in the case of a negative coefficient and white stands for a non-inclusion (a zero coefficient). The width of the column is proportional to the model's posterior probability. You can see in figure 1 that `complaints` is included in every model mass with a negative coefficient, and `raises` for example is included two times only and with different coefficient signs.
Now I will go on with using the R package BMS [Feldkircher and Zeugner, 2015b] to perform the BMA approach. Therefore, I have to load the BMS package first:

```
R> library("BMS")
```

and perform BMA via the function `bms()` and write results into the variable `att`:

```
R> att<-bms(attitude)
```

The `bms()` function samples all possible model combinations via the so-called Birth-death sampler (MC3, "bd") or enumeration and returns aggregate results. [Feldkircher and Zeugner, 2015b] The default prior used in function `bms()` is the binomial-beta model prior with random θ. Regarding the MCMC sampler the default option of the function is the enumeration of all models, which, instead of an approximation by means of any MCMC sampling scheme, evaluates all possible models. Only if this enumeration becomes time-consuming or infeasible for many variables the MC3 sampler will be used as a MCMC sampler, which is the standard model sampler used in most BMA routines. The default number of burn-in draws for the MC3 sampler is 1000. For more details about the theoretical background of this package and about its functions I refer to the paper of Zeugner Stefan and Feldkircher Martin [Feldkircher and Zeugner, 2015a].
The output of `bms(attitude)` you can find in table 2.
And some details about the BMA approach:

```
Mean no. regressors: 1.7773, Draws: 64, Burnins: 0, Time: 0.1060739 secs,
```

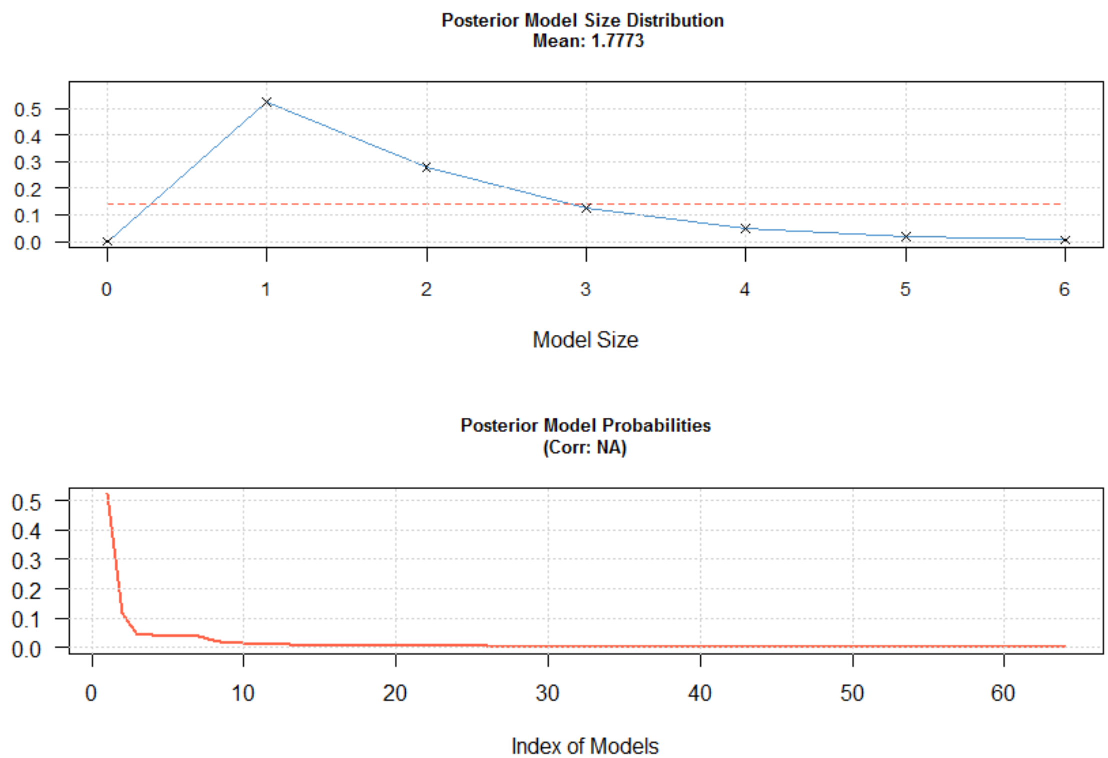
Table 2: Attitude dataset, BMS package, bms(attitude)

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
complaints	1.00	0.70	0.12	1.00	1.00
learning	0.27	0.07	0.13	1.00	3.00
advance	0.15	-0.02	0.08	0.00	6.00
privileges	0.12	-0.01	0.05	0.00	2.00
raises	0.12	0.01	0.07	0.77	4.00
critical	0.11	0.00	0.05	0.91	5.00

`No. models visited: 64, Modelspace` 2^K`: 64, % visited: 100, % Topmodels: 100,`
`Corr PMP: NA, No. Obs.: 30, Model Prior: random / 3, g-Prior: UIP,`
`Shrinkage-Stats: Av=0.9677`

In figure 2 you find a plot with the Posterior Model Size Distribution and the Posterior Model Probabilities.

Figure 2: BMS package, attitude dataset



The matrix (table 2) shows the variable names and corresponding statistics: Column Post Mean shows the coefficients averaged over all models (including the models wherein the variable was not contained). The covariate `complaints` has a comparatively large coefficient and seems to be most important. Column PIP represents the posterior inclusion probabilities which shows the importance of the variables in explaining the data. (the sum of posterior model probabilities (PMPs) for all models wherein a covariate was included) We see that with 99.96% most of posterior model mass rests on models that include `complaints`, contrarily `learning` has an intermediate PIP of 27.4% and all residual covariates do not seem to matter much. Column

Post SD (posterior standard deviations) provides that `complaints` is certainly positive, `advance` most likely negative and the corresponding number of `privileges` is near to zero.
The 5 topmodels can be received with using:

```
R> topmodels.bma(att)[,1:5]
```

and with the output in table 3.

Table 3: BMS package, attitude dataset, 5 topmodels

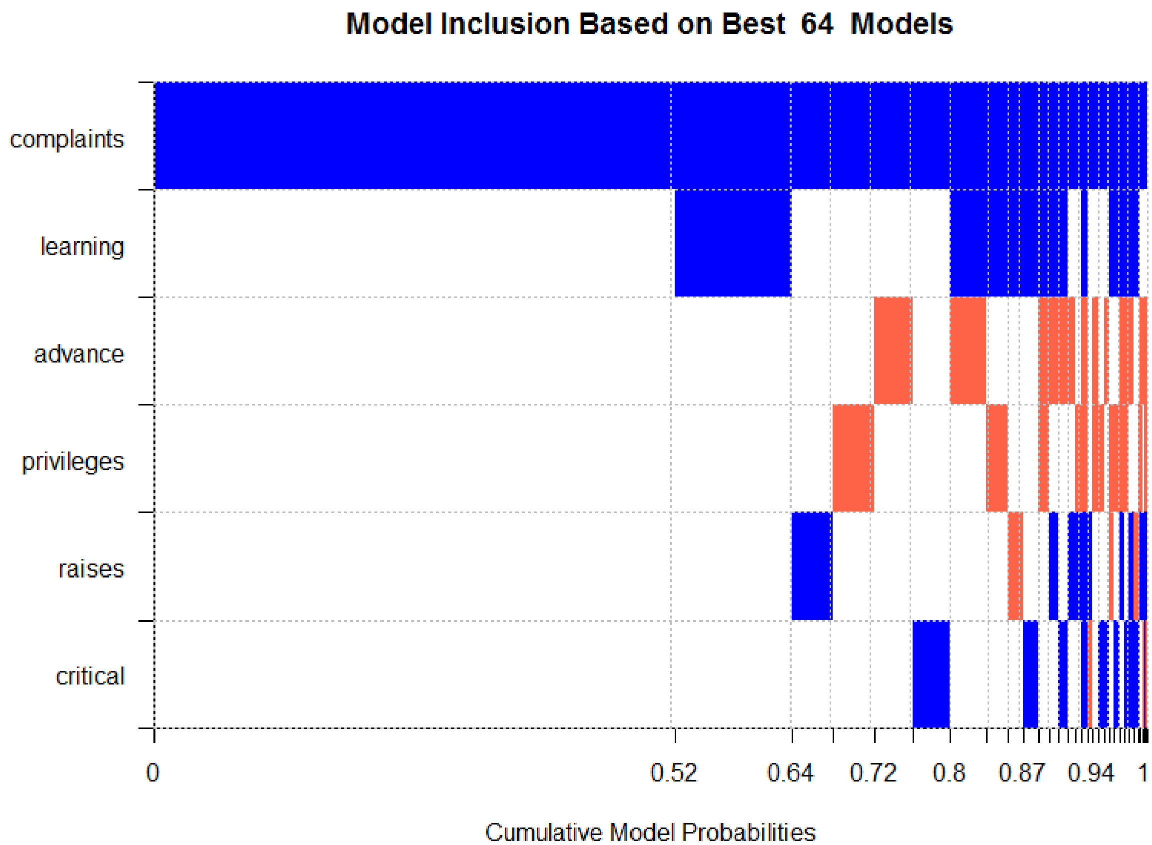
	20	28	24	30	21
complaints	1.00	1.00	1.00	1.00	1.00
privileges	0.00	0.00	0.00	1.00	0.00
learning	0.00	1.00	0.00	0.00	0.00
raises	0.00	0.00	1.00	0.00	0.00
critical	0.00	0.00	0.00	0.00	0.00
advance	0.00	0.00	0.00	0.00	1.00
PMP (Exact)	0.52	0.12	0.04	0.04	0.04
PMP (MCMC)	0.52	0.12	0.04	0.04	0.04

which is a binary matrix with the variables arranged row-wise and the models column-wise. For a particular model a 0 indicates exclusion and 1 inclusion of a variable associated with a given row.
To get a more comprehensive overview of the models, I use the command

```
R> image(att)
```

that produces figure 3.

Figure 3: BMS package, attitude dataset



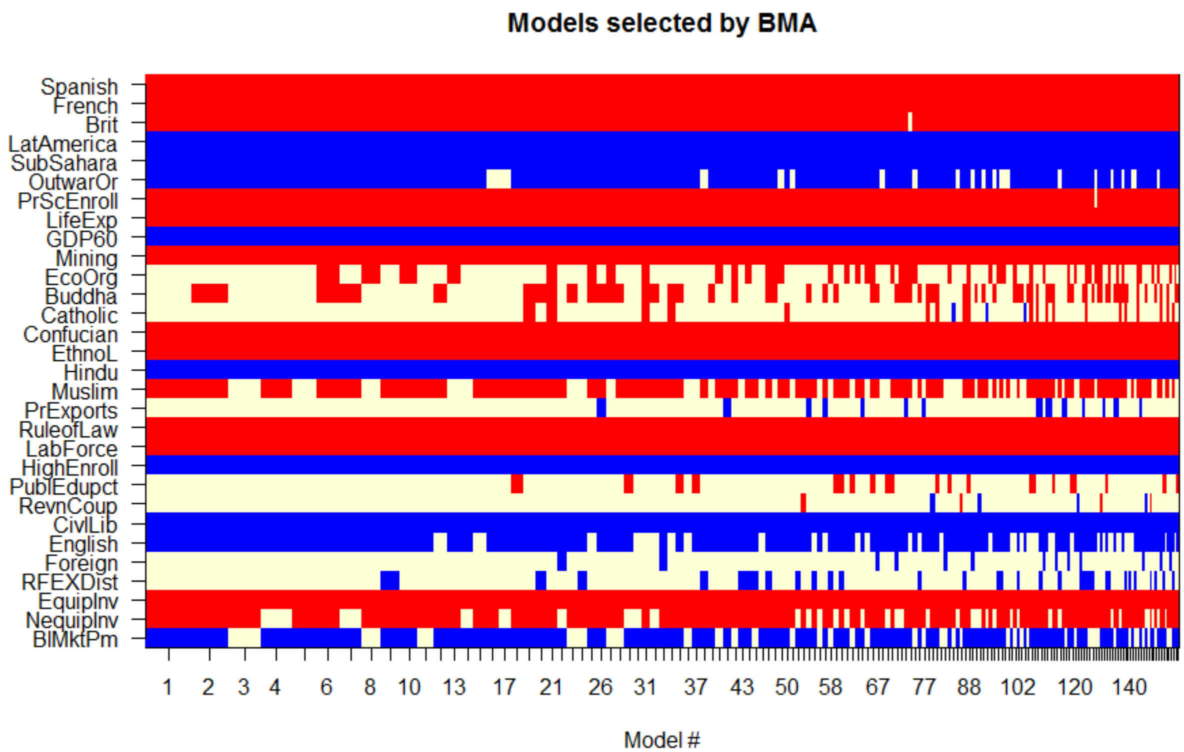
In figure 3 blue color corresponds to a positive and red to a negative and white to a zero coefficient (non-inclusion). The horizontal axis shows the best models scaled by their PMPs, where you can see that the best model with most mass only includes `complaints` with a posterior model probability of 52% and that `complaints` is included in every model mass with a positive coefficient. In contrast `raises` is included infrequently and its coefficient sign changes according to the model.
The main differences of the results of the two packages, regarding the BMA approach on the attitude dataset, are mostly due to the explained differences in the default parameters in the two functions used.
As you can see in figure 1 and 3 in both packages variable `complaints` is included in nearly every model, but in the BMA package with a negative coefficient and in the BMS package with a negative coefficient. In both packages the second best model additionally includes variable `learning`, but again both packages with a different coefficient sign, BMA a negative and BMS a positive.

6.2 FLS dataset

The second dataset which I will use to show BMA approaches is the dataset of FLS, with 41 explanatory variables of 72 countries. This dataset is available in package BMS. And again I want to show the differences in running BMA approaches between the two packages BMA and BMS.
Again we perform BMA via the function `bicreg()` and write the results into the variable `fls` and take a look at the coefficient results with the output in table 4.

The output in table 4 shows that variables like `Spanish, French, Brit, LatAmerica, SubSahara, OutwarOr, PrScEnroll, LifeExp, GDP60, Mining, Confucian, EthnoL, Hindu, RuleofLaw, LabForce, HighEnroll, CivlLib` and `EquipInv` have a posterior probability of being in the model of over 90%. Variables as `Muslim, English, NequipInv` and `BlMktPm` have also a high posterior probability but all others a low one.
And again to get a more comprehensive overview of the models I produce figure 4 with the command `imageplot.bma(fls)`.

Figure 4: FLS dataset, BMA package

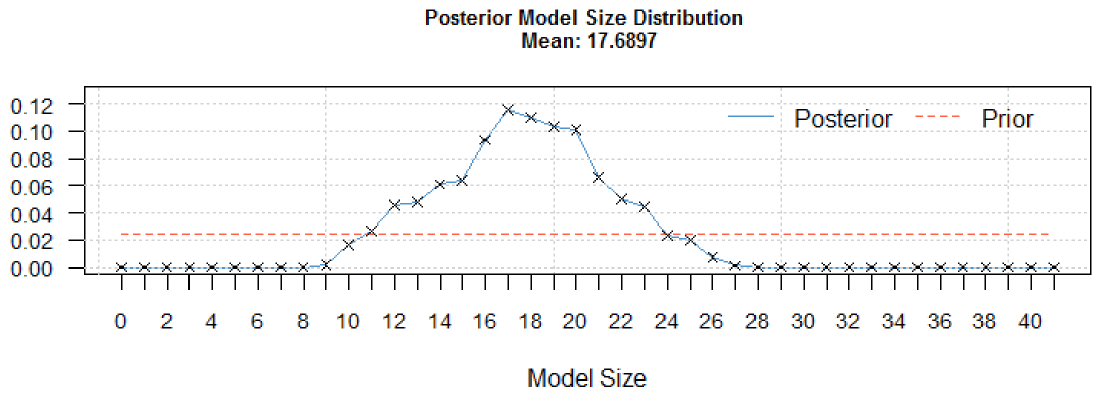


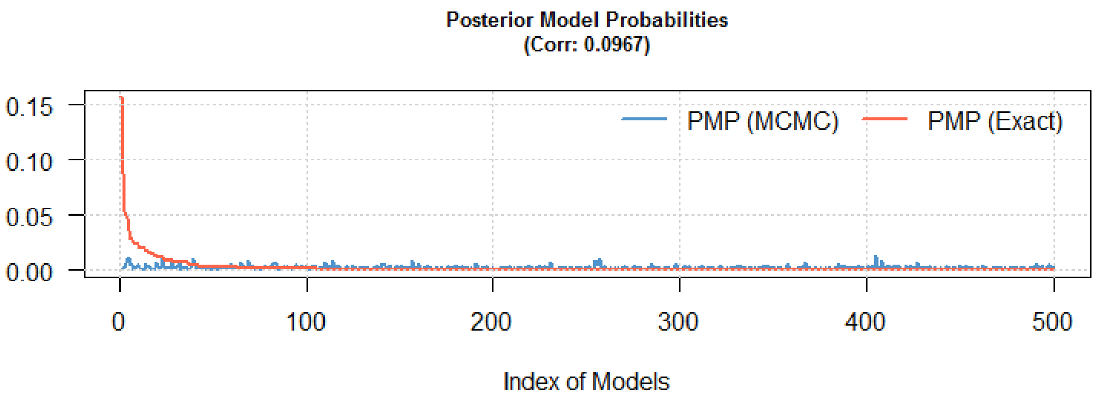
When using the BMS package and function `bms()` I get the output of table 5.
And some details about the BMA approach:

```
Mean no. regressors: 17.6897, Draws: 3000, Burnins: 1000, Time: 2.48877 secs,
No. models visited: 1217, Modelspace 2^K: 2.2e+12, % visited: 5.5e-08,
% Topmodels: 64, Corr PMP: 0.0967, No. Obs.: 72, Model Prior: random / 20.5,
g-Prior: UIP, Shrinkage-Stats: Av=0.9863
```

Figure 5 shows the Posterior Model Size Distribution and the Posterior Model Probabilities.

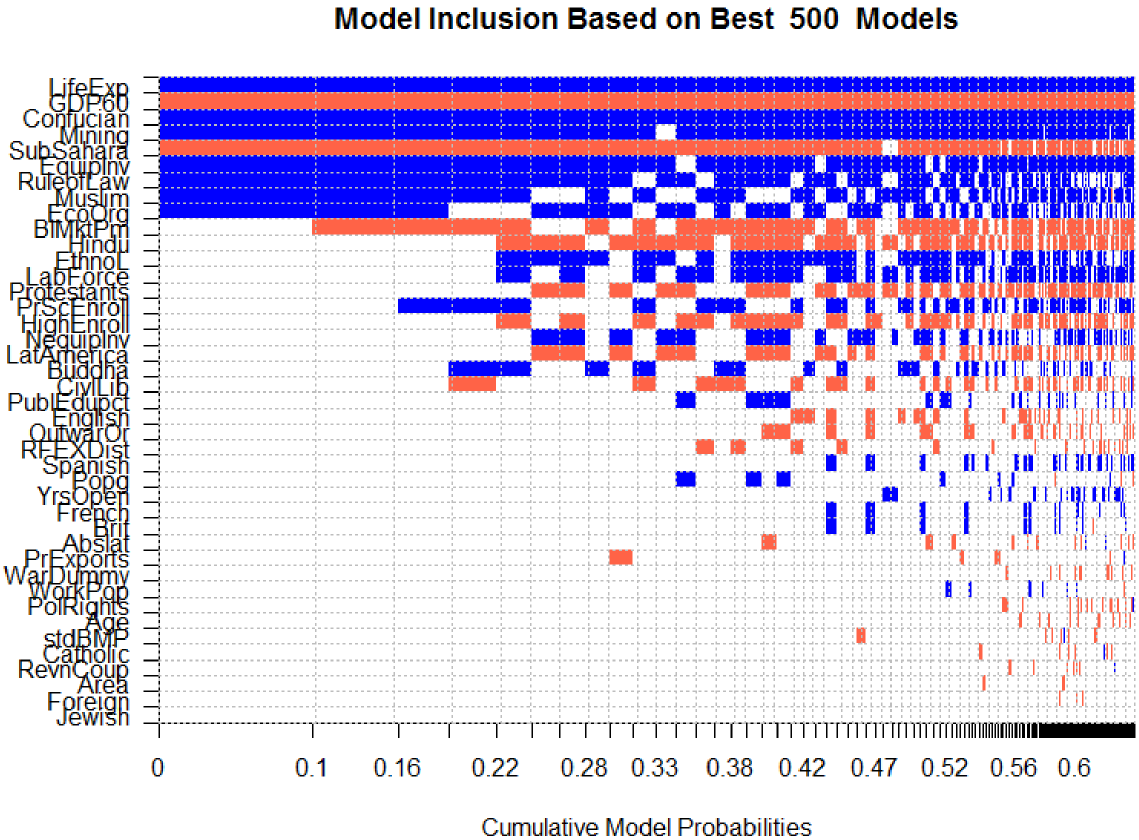
Figure 5: FLS dataset,BMS package





The matrix (table 5) shows the variable names and corresponding statistics: Column Post Mean shows the coefficients averaged over all models (including the models wherein the variable was not contained). The covariate `EquipInv` has a comparatively large coefficient and seems to be most important. Regarding the importance of the variables in explaining the data, described by PIP, you can see that with 100% all of posterior model mass rests on models that include `LifeExp`, `GdP60` and `Confucian`. Additionally covariates like `EquipInv`, `Mining` and `SubSahara` have also an intermediate PIP of over 80%, while other variables have low posterior probabilities.
The 5 topmodels can be found in table 6.
And a more comprehensive overview of the models in figure 6.

Figure 6: FLS dataset, BMS package



In figure 6 you can see that the best model with most mass include the covariates `LifeExp`, `GDP60`, `Confucian`, `Mining`, `SubSahara`, `EquipInv`, `RuleofLaw`, `Muslim` and `EcoOrg` with a posterior model probability of about 10% and that `LifeExp`, `GDP60` and `Confucian` are included in every model mass, `LifeExp` and `Confucian` with a positive coefficient and `GDP60` with a negative coefficient.

As you can see in figure 4 and 6 nearly all of the most important variables found in the BMS package are also important when using the BMA package. Again, as in section 6.1 already seen, coefficient signs of some variables are different. In general it seems that the BMA package can not really afford usable results in using the FLS dataset. In contrast, the BMS package, with its results in figure 6, seems to identify the real important variables. Therefore, I would recommend using the BMS package for this dataset.

6.3 Comparison of both packages

Amini and Parameter [Amini and Parmeter, 2012] conclude that package BMS is the only package among its competitors that is able to reproduce empirical results in FLS and MP. It is also very flexible regarding the use of prior information and allows different priors. It provides various specifications for Zellner's g prior and comes along with numerous graphical tools to analyze posterior coefficient densities, the posterior model size or predictive densities. The flexibility and user-friendliness of this package is proved by the usage of package BMS in lots of studies performing Bayesian Averaging Models.[Feldkircher and Zeugner, 2015a]

7 Conclusion

As we have seen in the literature, researchers typically acting as if the chosen model is the true model and therefore tend to produce optimistic conclusions due to the underestimation of the uncertainty associated with the whole estimation. In the last two to three decades, researchers made a change regarding the uncertainty surrounding the selection of empirical models and more and more try to solve this challenge. Model averaging approaches for example estimate the effect of interest under all possible combinations of controls and report a weighted average effect of control explanatory variables.
This paper presented a theoretical overview of Bayesian statistics, Model Selection and especially about Bayesian Model Averaging. It gives an overview to the historical changes in the last years regarding the BMA approach and shows its application in empirical growth.
Following the presented literature many different problems arise in BMA approaches and are extremely critical for the outcomes of its analysis. Especially prior assumptions clearly matter and LSE12 therefore recommend two priors, the benchmark Beta prior and the hyper-g/n prior, that stand out by not having any truly bad behavior. Another potential problem is the multicollinearity among the set of potential explanatory variables, where HEAs' concluding results indicate that Bayesian LASSO, Bayesian bridge regression and Bayesian elastic nets present better out-of-sample prediction properties than standard model averaging methods. Also different jointness measures have been presented by LSM7, LSM9 and CEA16, which used LCA to measure it.
For computing BMA approaches I would recommend the R package BMS, which is very flexible regarding the use of prior information, comes along with many graphical tools, is flexible and user-friendly.

References

[Amini and Parmeter, 2012] Amini, S. and Parmeter, C. (2012). Coparisons of model averaging techniques: Assessing growth determinants. *Journal of Applied Econometrics*, 27(5):870–876.

[Avramov, 2002] Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64:423–458.

[Brock and Durlauf, 2001] Brock, W. and Durlauf, S. (2001). Growth empirics and reality. *World Bank Economic Review*, 15:229–272.

[Brock et al., 2003] Brock, W., Durlauf, S., and West, K. (2003). Policy evaluation in unvertain economic environments. *Brookings Papers on Economic Activity*, 1:235–322.

[Brock et al., 2006] Brock, W., Durlauf, S., and West, K. (2006). Model uncertainty and policy evaluation: some theory and empirics. *Journal of Econometrics*, 136:629–664.

[Chipman, 1996] Chipman, H. (1996). Bayesian variable selection with related predictors. *Canadian Journal of Statistics*, 24:17–36.

[Clyde and George, 2000] Clyde, M. and George, E. (2000). Flexible empirical bayes estimation for wavelets. *Journal of the Royal Statistics Society Series B*, 62:681–698.

[Cremers, 2002] Cremers, K. (2002). Stock return predictability: a bayesian model selection perspective. *The Review of Financial Studies*, 15:1223–1249.

[Crespo Cuaresma, 2010] Crespo Cuaresma, J. (2010). How different is africa? a comment on masanjala and papageorgiou. *Journal of Applied Econometrics*, 26(6):1041–1047.

[Crespo Cuaresma et al., 2016] Crespo Cuaresma, J., Hofmarcher, P., Humer, S., Moser, M., and Grün, B. (2016). Unveiling covariate inclusion structures in economic growth regressions using latent class analysis. *European Economic Review*, 81:189–202.

[Crespo Cuaresma and Slacik, 2009] Crespo Cuaresma, J. and Slacik, T. (2009). On the determinants of currency crises: the role of model uncertainty. *Journal of Macroeconomics*, 31:621–632.

[Doppelhofer and Weeks, 2009] Doppelhofer, G. and Weeks, M. (2009). Jointness of growth determinants. *Journal of Applied Econometrics*, 24(2):209–244.

[Durlauf et al., 2008] Durlauf, S., Kourtellos, A., and Tan, C. (2008). Are any growth theories robust? *Economic journal*, 118:329–346.

[Edgerton and Kolbe, 1936] Edgerton, H. and Kolbe, L. (1936). The method of minimum variation for the combination of criteria. *Psychometrika*, 1:183–188.

[Feldkircher and Zeugner, 2015a] Feldkircher, M. and Zeugner, S. (2015a). Bayesian model averaging employing fixed and flexible priors: The bms package for r. *Journal of Statistical Software*, 68(4).

[Feldkircher and Zeugner, 2015b] Feldkircher, M. and Zeugner, S. (2015b). Bms: Bayesian model averaging library. `https://cran.r-project.org/web/packages/BMS/index.html`. Accessed: 2017-13-04.

[Fernández et al., 2001a] Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark priors for bayesian model averaging. *Journal of Econometrics*, 100:381–427.

[Fernández et al., 2001b] Fernández, C., Ley, E., and Steel, M. F. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16:563–576.

[Foster and George, 1994] Foster, D. and George, E. (1994). The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22:1947–1975.

[Garratt et al., 2009] Garratt, A., Koop, G., Mise, E., and Vahey, S. (2009). Real-time prediction with u.k. monetary aggregates in the presence of model uncertainty. *Journal of Business and Economic Statistics*, 27:480–491.

[Geisel, 1973] Geisel, M. (1973). Bayesian comparisons of simple macroeconomic models. *Journal of Money, Credit and Banking*, 5:751–772.

[Geisser, 1965] Geisser, S. (1965). A bayes approach for combining correlated estimates. *Journal of the American Statistical Association*, 60:602–607.

[George, 1999] George, E. (1999). Discussion of 'bayesian model averaging and model search strategies' by m. clyde. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 6*. Oxford University Press.

[George, 2010] George, E. I. (2010). Dilution priors: Compensating for model space redundancy. *IMS Collections Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown*, 6:158–165.

[George and Foster, 2000] George, E. I. and Foster, D. P. (2000). Calibration and empirical bayes variable selection. *Biometrika*, 87:731–747.

[Halperin, 1961] Halperin, M. (1961). Obtaining a composite measure from a number of different measures of the same attribute. *Journal of the American Statistical Association*, 56:36–43.

[Hansen and Yu, 2001] Hansen, M. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96:746–774.

[Held and Bové, 2014] Held, L. and Bové, D. S. (2014). *Applied Statistical Inference*. Springer-Verlag Berlin Heidelberg.

[Hoeting et al., 1999] Hoeting, A., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417.

[Hofmarcher et al., 2015] Hofmarcher, P., Crespo Cuaresma, J., Grün, B., and Hornik, K. (2015). Last night a shrinkage saved my life: economic growth, model uncertainty and correlated regressors. *Journal of Forecasting*, 34(2):133–144.

[Horst, 1938] Horst, P. (1938). Obtaining a composite measure from a number of different measures of the same attribute. *Psychometrika*, 1:53–60.

[Kass and Wasserman, 1995] Kass, R. and Wasserman, L. (1995). A reference bayesian test for nested hypothesis with large samples. *Journal of the American Statistical Association*, 90:928–934.

[Koop and Potter, 2004] Koop, G. and Potter, S. (2004). Forecasting in dynamic factor models using bayesian model averaging. *The Econometrics Journal*, 7:550–565.

[Koske and Wanner, 2013] Koske, I. and Wanner, I. (2013). The drivers of labour income inequality: an analysis based on bayesian model averaging. *Applied Economics Letters*, 20:123–126.

[Laplace, 1818] Laplace, P. (1818). *Deuxime Supplément a la Théorie Analytique des Probabilités*. paris: courcier.

[Leamer, 1978] Leamer, E. (1978). *Specification Searches*. new york: john wiley and sons.

[Leamer, 1983] Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73:31–43.

[Lempers, 1971] Lempers, F. (1971). *Psterior Probabilities of Alternative Linear Models*. Rotterdam: University Press.

[Levine and Renelt, 1992] Levine, R. and Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *The American Economic Review*, 82(4):942–963.

[Ley and Steel, 2007] Ley, E. and Steel, M. F. (2007). Jointness in bayesian variable selection with applications to growth regression. *Journal of Macroeconomics*, 29:476–493.

[Ley and Steel, 2009] Ley, E. and Steel, M. F. (2009). On the effect of prior assumptions in bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24:651–674.

[Ley and Steel, 2012] Ley, E. and Steel, M. F. (2012). Mixtures of g-priors for bayesian model averaging with economic applications. *Journal of Econometrics*, 171(2):251–266.

[Liang et al., 2008] Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103:410–423.

[Madigan and Raftery, 1994] Madigan, D. and Raftery, A. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89:1535–1546.

[Madigan and York, 1995] Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63:215–232.

[Masanjala and Papageorgiou, 2008] Masanjala, W. H. and Papageorgiou, C. (2008). Rough and lonely road to prosperity: a reexamination of the sources of growth in africa using bayesian model averaging. *Journal of Applied Econometrics*, 23:671–682.

[Mitchell et al., 2011] Mitchell, J., Pain, N., and Riley, R. (2011). The drivers ofinternational migration to the uk: a panel-based bayesian model averaging approach. *Economic Journal*, 121:1398–1444.

[Moral-Benito, 2015] Moral-Benito, E. (2015). Model averaging in economics: An overview. *Journal of Economic Surveys*, 29:46–75.

[Morales et al., 2006] Morales, K., Ibrahim, J., Chen, C., and Ryan, L. (2006). Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *Journal of the American Statistical Association*, 101:9–17.

[Moulton, 1991] Moulton, B. (1991). A bayesian approach to model selection and estimation with application to price indexes. *Journal of Econometrics*, 49:169–193.

[Pesaran et al., 2009] Pesaran, H., Schleicher, C., and Zaffaroni, P. (2009). Model averaging in risk management with an application to futures markets. *Journal of Empirical Finance*, 16:280–305.

[Raftery et al., 2015] Raftery, A., Hoeting, J., Colinsky, C., Painter, I., and Yeung, K. (2015). Bma: Bayesian model averaging. `https://cran.r-project.org/web/packages/BMA/index.html`. Accessed: 2017-13-04.

[Roberts, 1965] Roberts, H. (1965). Probabilisitic prediction. *Journal of the American Statistical Association*, 60:50–62.

[Sala-i Martin, 1992] Sala-i Martin, X. (1992). I just ran 2 million regressions. *American Economic Review*, 87:178–183.

[Sala-i Martin et al., 2004] 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.

[Spiegelhalter and Smith, 1982] Spiegelhalter, D. and Smith, A. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistics Society Series B*, 44:377–387.

[Vermunt and Magidson, 2002] Vermunt, J. and Magidson, J. (2002). Latent class cluster analysis. In Hagenaars, J. and McCutcheon, A., editors, *Applied Latent Class Analysis*, pages 89–106. Cambridge University Press.

[Wright, 2008a] Wright, J. (2008a). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146:329–341.

[Wright, 2008b] Wright, J. (2008b). Forecasting us inflation by bayesian model averaging. *Journal of Forecasting*, 28:131–144.

Table 4: FLS dataset, BMA package, bicreg(fls)

	p!=0	EV	SD	model 1	model 2	model 3	model 4	model 5
Intercept	100.0	7.706e-02	1.696e-02	7.965e-02	7.333e-02	7.872e-02	7.762e-02	8.286e-02
Spanish	100.0	1.431e-02	3.959e-03	1.399e-02	1.344e-02	1.626e-02	1.324e-02	1.551e-02
French	100.0	1.121e-02	3.198e-03	1.096e-02	1.045e-02	1.380e-02	1.042e-02	1.287e-02
Brit	99.6	7.617e-03	2.684e-03	7.855e-03	7.366e-03	1.014e-02	7.686e-03	9.296e-03
LatAmerica	100.0	-1.340e-02	4.384e-03	-1.270e-02	-1.162e-02	-1.611e-02	-1.261e-02	-1.669e-02
SubSahara	100.0	-2.187e-02	4.270e-03	-2.180e-02	-1.998e-02	-2.537e-02	-2.159e-02	-2.594e-02
OutwarOr	90.5	-3.440e-03	1.914e-03	-3.451e-03	-3.893e-03	-3.890e-03	-3.347e-03	-3.733e-03
PrScEnroll	99.7	2.228e-02	7.115e-03	2.487e-02	2.416e-02	1.897e-02	2.797e-02	2.041e-02
LifeExp	100.0	8.906e-04	1.848e-04	8.959e-04	8.931e-04	8.602e-04	9.110e-04	8.420e-04
GDP60	100.0	-1.815e-02	2.146e-03	-1.875e-02	-1.783e-02	-1.785e-02	-1.844e-02	-1.809e-02
Mining	100.0	3.340e-02	1.106e-02	3.277e-02	3.439e-02	3.046e-02	3.588e-02	3.365e-02
EcoOrg	26.6	2.675e-04	5.738e-04	.	.	.	.	.
Buddha	39.2	2.796e-03	4.471e-03	.	6.365e-03	.	.	.
Catholic	9.9	3.257e-04	1.451e-03	.	.	.	.	.
Confucian	100.0	7.444e-02	9.949e-03	7.593e-02	7.669e-02	7.200e-02	7.496e-02	7.219e-02
EthnoL	100.0	1.602e-02	3.710e-03	1.654e-02	1.604e-02	1.656e-02	1.708e-02	1.623e-02
Hindu	100.0	-1.090e-01	1.881e-02	-1.108e-01	-1.065e-01	-1.240e-01	-1.067e-01	-1.202e-01
Muslim	70.4	6.125e-03	5.365e-03	7.757e-03	8.898e-03	.	8.375e-03	.
PrExports	7.2	-2.784e-04	1.579e-03	.	.	.	.	.
RuleofLaw	100.0	1.261e-02	3.690e-03	1.313e-02	1.234e-02	1.351e-02	1.306e-02	1.277e-02
LabForce	100.0	3.720e-07	7.092e-08	3.797e-07	3.678e-07	4.044e-07	3.741e-07	3.974e-07
HighEnroll	100.0	-1.246e-01	2.828e-02	-1.213e-01	-1.177e-01	-1.330e-01	-1.276e-01	-1.306e-01
PublEdupct	10.2	9.389e-03	3.982e-02	.	.	.	.	.
RevnCoup	2.7	2.637e-06	5.857e-04	.	.	.	.	.
CivlLib	100.0	-2.632e-03	8.131e-04	-2.841e-03	-2.832e-03	-2.665e-03	-2.716e-03	-2.544e-03
English	82.1	-6.400e-03	4.462e-03	-7.787e-03	-7.226e-03	-8.379e-03	-8.027e-03	-8.139e-03
Foreign	5.4	-9.082e-05	6.795e-04	.	.	.	.	.
RFEXDist	15.4	-4.024e-06	1.291e-05	.	.	.	.	.
EquipInv	100.0	1.517e-01	3.427e-02	1.511e-01	1.475e-01	1.530e-01	1.649e-01	1.462e-01
NequipInv	79.0	2.793e-02	2.073e-02	2.946e-02	2.936e-02	3.816e-02	.	3.202e-02
BlMktPm	76.7	-4.426e-03	3.418e-03	-5.513e-03	-5.657e-03	.	-6.732e-03	-4.196e-03
nVar				22	23	20	21	21
r2				0.947	0.949	0.939	0.943	0.942
BIC				-1.170e+02	-1.165e+02	-1.162e+02	-1.161e+02	-1.157e+02
post prob				0.045	0.035	0.032	0.030	0.024

Table 5: FLS dataset, BMS package, bms(fls)

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
GDP60	1.00	-0.02	0.00	0.00	12.00
Confucian	1.00	0.06	0.02	1.00	19.00
EquipInv	0.96	0.15	0.06	1.00	38.00
LifeExp	0.90	0.00	0.00	1.00	11.00
RuleofLaw	0.80	0.01	0.01	1.00	26.00
SubSahara	0.80	-0.01	0.01	0.00	7.00
Muslim	0.80	0.01	0.01	0.99	23.00
PrScEnroll	0.75	0.02	0.01	0.99	10.00
Mining	0.64	0.02	0.02	1.00	13.00
BlMktPm	0.60	-0.00	0.00	0.00	41.00
Hindu	0.56	-0.04	0.05	0.00	21.00
EthnoL	0.51	0.01	0.01	1.00	20.00
CivlLib	0.50	-0.00	0.00	0.00	34.00
NequipInv	0.49	0.02	0.03	1.00	39.00
LabForce	0.47	0.00	0.00	0.99	29.00
HighEnroll	0.45	-0.05	0.06	0.00	30.00
Buddha	0.42	0.00	0.01	1.00	17.00
Spanish	0.41	0.00	0.01	0.96	2.00
EcoOrg	0.40	0.00	0.00	1.00	14.00
Protestants	0.37	-0.00	0.01	0.03	25.00
French	0.33	0.00	0.00	0.94	3.00
Catholic	0.32	-0.00	0.00	0.59	18.00
LatAmerica	0.27	-0.00	0.01	0.00	6.00
Brit	0.26	0.00	0.00	0.90	4.00
OutwarOr	0.26	-0.00	0.00	0.00	8.00
YrsOpen	0.24	0.00	0.01	0.97	15.00
English	0.23	-0.00	0.00	0.00	35.00
PrExports	0.20	-0.00	0.00	0.05	24.00
WarDummy	0.18	-0.00	0.00	0.01	5.00
RFEXDist	0.17	-0.00	0.00	0.00	37.00
Age	0.16	-0.00	0.00	0.00	16.00
PolRights	0.16	-0.00	0.00	0.09	33.00
PublEdupct	0.15	0.01	0.05	0.72	31.00
RevnCoup	0.14	-0.00	0.00	0.44	32.00
Foreign	0.13	-0.00	0.00	0.30	36.00
Abslat	0.12	0.00	0.00	0.44	1.00
Popg	0.11	-0.00	0.07	0.32	27.00
stdBMP	0.10	-0.00	0.00	0.25	40.00
Jewish	0.09	0.00	0.00	0.81	22.00
WorkPop	0.06	-0.00	0.00	0.20	28.00
Area	0.05	-0.00	0.00	0.39	9.00

Table 6: FLS dataset, BMS package, 5 topmodels

	0046845800c	0046844800c	00468c1800c	0046845800d	004f0749889
Abslat	0.00	0.00	0.00	0.00	0.00
Spanish	0.00	0.00	0.00	0.00	0.00
French	0.00	0.00	0.00	0.00	0.00
Brit	0.00	0.00	0.00	0.00	0.00
WarDummy	0.00	0.00	0.00	0.00	0.00
LatAmerica	0.00	0.00	0.00	0.00	0.00
SubSahara	1.00	1.00	1.00	1.00	1.00
OutwarOr	0.00	0.00	0.00	0.00	0.00
Area	0.00	0.00	0.00	0.00	0.00
PrScEnroll	0.00	0.00	0.00	0.00	1.00
LifeExp	1.00	1.00	1.00	1.00	1.00
GDP60	1.00	1.00	1.00	1.00	1.00
Mining	0.00	0.00	0.00	0.00	1.00
EcoOrg	1.00	1.00	1.00	1.00	0.00
YrsOpen	0.00	0.00	0.00	0.00	0.00
Age	0.00	0.00	0.00	0.00	0.00
Buddha	0.00	0.00	0.00	0.00	0.00
Catholic	0.00	0.00	1.00	0.00	0.00
Confucian	1.00	1.00	1.00	1.00	1.00
EthnoL	0.00	0.00	0.00	0.00	1.00
Hindu	0.00	0.00	0.00	0.00	1.00
Jewish	0.00	0.00	0.00	0.00	0.00
Muslim	1.00	1.00	0.00	1.00	1.00
PrExports	0.00	0.00	0.00	0.00	0.00
Protestants	1.00	0.00	1.00	1.00	0.00
RuleofLaw	1.00	1.00	1.00	1.00	1.00
Popg	0.00	0.00	0.00	0.00	0.00
WorkPop	0.00	0.00	0.00	0.00	0.00
LabForce	0.00	0.00	0.00	0.00	1.00
HighEnroll	0.00	0.00	0.00	0.00	1.00
PublEdupct	0.00	0.00	0.00	0.00	0.00
RevnCoup	0.00	0.00	0.00	0.00	0.00
PolRights	0.00	0.00	0.00	0.00	0.00
CivlLib	0.00	0.00	0.00	0.00	1.00
English	0.00	0.00	0.00	0.00	0.00
Foreign	0.00	0.00	0.00	0.00	0.00
RFEXDist	0.00	0.00	0.00	0.00	0.00
EquipInv	1.00	1.00	1.00	1.00	1.00
NequipInv	1.00	1.00	1.00	1.00	0.00
stdBMP	0.00	0.00	0.00	0.00	0.00
BlMktPm	0.00	0.00	0.00	1.00	1.00
PMP (Exact)	0.14	0.11	0.05	0.04	0.03
PMP (MCMC)	0.01	0.00	0.00	0.00	0.00