EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1234 / AUGUST 2010

IMPOSING PARSIMONY IN CROSS-COUNTRY GROWTH REGRESSIONS

by Marek Jarociński

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1234 / AUGUST 2010

IMPOSING PARSIMONY IN CROSS-COUNTRY GROWTH REGRESSIONS [1]

by Marek Jarociński [2]

In 2010 all ECB publications feature a motif taken from the €500 banknote.

NOTE: This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the author and do not necessarily reflect those of the ECB.

[1] I am grateful to Antonio Ciccone and Domenico Giannone for useful conversations.

[2] European Central Bank, DG-Research, Kaiserstrasse 29,
D-60311 Frankfurt am Main, Germany;
e-mail: Marek.Jarocinski@ecb.europa.eu

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

The number of variables related to long-run economic growth is large compared with the number of countries. Bayesian model averaging is often used to impose parsimony in the cross-country growth regression. The underlying prior is that many of the considered variables need to be excluded from the model. This paper, instead, advocates priors that impose parsimony without excluding variables. The resulting models fit the data better and are more robust to revisions of income data. The positive relationship between measures of trade openness and growth is much stronger than found in the literature.

JEL classification: C20, C52, O40, O47

Keywords: Economic Growth, Bayesian Model Averaging, Adaptive Ridge Regression, Measurement Error

Non-technical summary

This paper proposes a new econometric approach to studying determinants of economic growth across countries. It estimates a large regression in which growth is regressed on all available explanatory variables that have been proposed in the literature. The Introduction argues that this approach is consistent with the existing growth theory. The rest of the paper shows that it is not only feasible but also empirically plausible. Moreover, it delivers results which are robust to the measurement error inherent in the data.

Since the number of coefficients is large, it is necessary to impose some parsimony on the estimation. This paper compares a range of approaches to achieve this. It shows advantages of a simple approach that has not received enough attention before: shrinking the coefficients towards zero with variants of the well known *ridge regression*.

The number of explanatory variables is so large because hundreds of theories have been developed to explain economic growth. The literature review of Durlauf, Johnson, Temple (2005) finds as many as 145 different variables included in growth regressions in published papers. However, their relationship with long-run growth (over a horizon of 30 years or more) can usually be studied using a sample of at most 100 country observations. An additional, crucial fact is the "theory open-endedness" (pointed out by Brock and Durlauf 2001 and Durlauf, Johnson, Temple 2005): the observation that all these growth theories tend to be mutually compatible. Therefore, growth theory does not impose restrictions on the growth regression specification.

This paper interprets theory open-endedness as suggesting that all the variables suggested by the theories should be included simultaneously in the growth regression. Including all variables simultaneously is crucial to avoid omitted variables bias. Given the limited available data, one cannot perfectly control for all variables proposed in the literature. However, this paper makes an effort to control as well as possible. Controlling for all variables turns out to be very important for the estimated coefficients. Empirically, it turns out that although many variables matter little individually, they matter a lot when taken together.

Majority of empirical growth papers simply exclude many potentially relevant variables and these exclusion decisions are often ad hoc. In a seminal paper Levine and Renelt (1992) argue that most results of this literature are not robust to the choice of variables. Since Brock and Durlauf (2001), Fernandez, Ley and Steel (2001b) and Sala-i-Martin, Doppelhofer, Miller (2004) a large literature uses Bayesian model averaging (BMA). BMA involves estimating many growth regressions which only include few variables. Then results of interest are averaged across these small regression models. In this way, results are conditional on all these specifications simultaneously. BMA often performs very well in forecasting. However, in the context of growth regressions the focus is on the estimation of partial regression coefficients. All small models suffer from the omitted variables bias. It is not clear that averaging over such small models yields correct partial regression coefficients.

This paper compares a range of Bayesian priors. It uses a general framework which nests BMA, adaptive ridge and ridge models as special cases. The empirical part of this paper studies the well-known dataset of Sala-i-Martin, Doppelhofer and Miller (2004). A number of interesting empirical results are found:

First, the studied ridge-type models tend to fit the data better than standard specifications of BMA. The discussion above questions the a priori appeal of the exclusion restrictions in the BMA. Now it turns out that also a posteriori they are not attractive as they do not yield superior fit.

Second, ridge-type models are much more robust to revisions of the growth data. This is important because Ciccone and Jarociński (2010) show that many published BMA results are not consistent across vintages of the datasets. Data uncertainty is inherent in empirical growth and it is crucial to have a model which is not excessively sensitive to it.

Third, this paper selects a baseline prior which delivers both a good fit and a high degree of robustness to the dataset vintage. This prior turns out to be very close to the standard ridge regression. With the baseline prior, the conditional convergence of income is much slower and the effect of Primary Schooling is half of that found using BMA. However, more coefficients are economically relevant than when BMA is used. Most interestingly, various measures of trade openness, which has been hotly debated in the literature, are found to be positively related to growth. This contrasts with the mixed but mostly negative evidence from BMA. Variables whose relation with growth is very weak regardless of the dataset vintage include Malaria Prevalence, Fertility, Population Density and the Fraction of Muslims. This contrasts with BMA results which, as shown in Ciccone and Jarociński (2010), include strong effects of these variables in some, but not all dataset vintages.

The good news is that the approach advocated in this paper is computationally very simple. The bottom line of the paper is that a robust analysis of a large cross-country dataset can be performed with a simple ridge regression, which is available in most econometric packages and involves only one matrix inversion. This allows empirical growth researchers to shift their attention from computational issues to the other challenges facing empirical growth research, such as endogeneity of growth determinants, nonlinearity and new data collection.

1 Introduction

This paper estimates a cross-country growth regression with many explanatory variables. Since the number of coefficients is large, it is necessary to impose some parsimony on the estimation. This paper compares a range of approaches to achieve this. It shows advantages of a simple approach that has not received enough attention before: shrinking the coefficients towards zero with variants of the well known ridge regression.

Imposing parsimony in a convincing way is crucial, because hundreds of theories have been developed to explain economic growth. The literature review of Durlauf et al. (2005) finds as many as 145 different variables included in growth regressions in published papers. However, their relationship with long-run growth (over a horizon of 30 years or more) can usually be studied using a sample of at most 100 country observations.[1] What complicates matters further is that, as pointed out by Brock and Durlauf (2001) and Durlauf et al. (2005), all these theories tend to be mutually compatible ("theory open-endedness").

This paper interprets theory open-endedness as suggesting that all the variables suggested by the theories should be included simultaneously in the growth regression. In coefficient estimation it is important to control for all other variables to avoid the omitted variables bias. One cannot do this perfectly in the available samples but it is worth going as far as possible. This paper finds that although many variables matter little individually, they matter a lot when taken together.

Most empirical growth papers simply exclude many potentially relevant variables and these exclusion decisions are often ad hoc. Levine and Renelt (1992) argue that most results of this literature are not robust to the choice of variables. Since Brock and Durlauf (2001), Fernández et al. (2001b) and Sala-i-Martin et al. (2004) a large literature uses Bayesian model averaging (BMA). BMA imposes parsimony by specifying a prior according to which every regression coefficient may be zero with a discrete probability, giving rise to different regression specifications.[2] BMA results are conditional on all these specifications. Specifications with subsets of explanatory variables may be attractive when forecasting is the goal. When several variables are

[1] Panel data can be used to increase the number of observations at the cost of reducing the horizon, which is often deemed by researchers to be undesirable. Also, Hauk and Wacziarg (2009) studies biases present in the panel estimation and argues for using a single cross-section of long-term growth observations.

[2] There are also frequentist approaches where implicit priors have this feature. They are applied to cross-country growth data eg, in Hendry and Krolzig (2004); Magnus et al. (2010); Wagner and Hlouskova (2009).

correlated it may be enough to include just one of them to forecast the dependent variable well.[3] However, when the interest is in partial regression coefficients, this body of small specifications is less attractive because they all suffer from omitted variables bias.

This paper compares a range of Bayesian priors. All these priors achieve parsimony by assuming a prior mean of zero for all coefficients. The priors differ in the prior variance. BMA results from one particular specification of the variance. Another specification of the variance leads to the adaptive ridge regression. By varying one of the prior hyperparameters, the adaptive ridge models cover a wide range, from setups very close to BMA at one extreme to the standard ridge regression at the other extreme.

The contribution of this paper is to state these alternative priors for growth regressions and to use them empirically. It seems that ridge-type priors have not been used for growth regressions before, although they have a long history. Ridge regression was introduced by Hoerl and Kennard (1970) to deal with multicollinearity in the data. Adaptive ridge regressions have been studied in many statistical papers since Strawderman (1978). Adaptive ridge regression replaces a discrete set of models in BMA with a continuous family of models, which nests that set. This is generally recommended whenever all the models in the continuous family also make scientific sense (as is the case here), see eg, Gelman et al. (2003, ch.15.5) or Sims (2003).

A number of interesting results emerge when a range of parsimony priors is applied to the well-known dataset of Sala-i-Martin et al. (2004).

First, the studied ridge-type models tend to fit the data better than standard specifications of BMA.[4] The discussion above questions the a priori appeal of the exclusion restrictions in BMA. A superior fit might justify such restrictions nevertheless. However, it turns out that they do not guarantee superior fit.

Second, ridge-type models tend to be much more robust to revisions of the growth data. This is important because Ciccone and Jarociński (2010) show that many published BMA results are not consistent across vintages of the datasets. Data uncertainty is inherent in empirical growth and it is crucial to have a model that is not excessively sensitive to it.

Third, the discussion focuses on a baseline prior that delivers both a good

[3] However, dropping variables does not necessarily lead to better forecasting models: De Mol et al. (2008) find that a ridge regression with all candidate variables forecasts as well as approaches that select variables or principal components, while its coefficients are much more stable. Denison and George (2001) find that adaptive ridge predicts better than BMA.

[4] This finding confirms and extends the results of Eicher et al. (2009) who find that BMA specifications with stronger shrinkage and larger prior model size fit the data better.

fit and a high degree of robustness to the dataset vintage. This prior turns out to be very close to the standard ridge regression. With this prior, the conditional convergence of income is much slower and the effect of Primary Schooling is half of that found using BMA. However, more coefficients are economically relevant than when BMA is used. Most interestingly, various measures of trade openness, which has been hotly debated in the literature, are found to be positively related to growth. This contrasts with the mixed but mostly negative evidence from BMA. Variables whose relation with growth is very weak regardless of the dataset vintage include Malaria Prevalence, Fertility, Population Density and the Fraction of Muslims. This contrasts with BMA results which, as shown in Ciccone and Jarociński (2010), include strong effects of these variables in some, but not all dataset vintages.

Another novelty of this paper is the focus on the economic significance of the estimated coefficients. The quoted literature, in contrast, places much weight on the statistical significance indicated by variables' inclusion probabilities.

Section 2 discusses the econometric specification of alternative parsimony priors. Section 3 describes the data and the model space. Section 4 reports the fit of the models and their robustness to data uncertainty. Section 5 reports the growth determinants found with the baseline specification, compares them with the findings of the BMA approach and performs an extensive sensitivity analysis. Section 6 concludes. Computational details and some additional results are reported in the Appendix.

2 Parsimony Priors for Linear Regressions

GDP growth (y) is related to K explanatory variables gathered in matrix X through the gaussian linear regression model:

$$y = \iota\alpha + X\beta + \varepsilon \qquad\qquad \varepsilon \sim \mathrm{N}\left(0, \sigma^2 I\right) \tag{1}$$

where the number of observations is N, ι is a vector of 1s, I is an $N \times N$ identity matrix and $(\alpha, \beta, \sigma^2)$ are unknown parameters. The explanatory variables in X are standardized (they have zero mean and unit standard deviation) to facilitate interpretation of the coefficients.

I use the usual noninformative priors for the constant term α and the error variance σ^2:

$$p(\alpha) \propto 1, \qquad p(\sigma^2) \propto \sigma^{-2} \tag{2}$$

It remains to specify a prior about β. While subjective priors about β may be available, they may also be contentious. Therefore, empirical growth researchers are also interested in lessons that can be drawn from the data using agnostic priors.

Unfortunately the usual noninformative prior (the flat prior $p(\beta) \propto 1$) is not a viable option here because of the large number of potential explanatory variables. When the number of variables (K) is large relative to the number of observations (N) then the data alone is not sufficient for reliable inference about β. One way to see this is to consider the flat-prior posterior, which is normally centered on the OLS estimate of β, $(X'X)^{-1}X'y$ and has a variance proportional to $(X'X)^{-1}$. When $K > N$ the $X'X$ matrix is not invertible. In this case the posterior mean does not exist and the variance is infinite. When K is smaller, but close to N, the matrix $X'X$ is badly conditioned. This implies that the posterior mean is very sensitive to small changes in the data and the posterior variance is hopelessly large.

To get more constructive results with large K we need to introduce some parsimony. The standard agnostic approach is to specify a prior for β that is centered on zero and thus constrains coefficients' absolute sizes. It is convenient to use the conjugate prior, which is gaussian, with the variance proportional to the error variance:

$$p(\beta) = N(0, \sigma^2 V) \tag{3}$$

Equations (1), (2) and (3) define the framework for the whole paper. The following subsections discuss alternative specifications of V in (3). I use two quantities to understand the effect of different assumptions on V: *shrinkage adaptivity* (defined later in this section) and *effective model size*.

The *effective model size* gives the effective number of estimated coefficients. A shrinkage prior reduces the effective model size because it restricts the coefficients. One exact restriction reduces by one the number of coefficients effectively estimated. The restrictions imposed in (3) are stochastic, not exact, and a given value of V implies the following effective model size:[5]

$$J = \operatorname{tr}\left(X(X'X + V^{-1})^{-1}X'\right) \tag{4}$$

2.1 Ridge Regression

The prior that gives rise to the ridge regression is:

$$p(\beta) = \mathrm{N}\left(0, \sigma^2 \operatorname{diag}(\tau)^{-1}\right) \tag{5}$$

[5] This expression is equal to the trace of the 'hat matrix' that projects y onto its fitted values. It is usual in linear models to take this quantity as the effective number of parameters. For an in-depth discussion of counting the effective parameters see eg, Hodges and Sargent (2001) or Spiegelhalter et al. (2002), who justify and generalize (4) in various ways. Notice that when the prior for β is noninformative and V^{-1} is a matrix of zeros we have that $J = K$ ie, the effective number of parameters equals the number of variables.

where diag(τ) denotes a matrix with τ on the main diagonal and zeros elsewhere, and $\tau = (\bar{\tau}, \ldots, \bar{\tau})$ is a vector with K constant elements $\bar{\tau}$. Parameter $\bar{\tau}$ determines shrinkage strength, which is fixed and common for all variables. Given X, the parameter $\bar{\tau}$ can be specified to deliver any desired effective model size.

2.2 Adaptive Ridge Regression

An adaptive ridge regression emerges when the shrinkage strength is adapted for each coefficient based on the data. In the present context, instead of making τ a fixed constant vector, I assume it is unknown and specify a prior for it. In the computation of the posterior the data update this prior. A posteriori good explanatory variables are shrunk more and poor explanatory variables are shrunk less.

The prior is that elements of τ, denoted $\tau_k, k = 1 \ldots K$ have independent and identical gamma densities with shape parameter $a > 0$ and inverse-scale parameter $b > 0$:

$$p(\tau_k) \propto \tau_k^{a-1} \exp(-b\tau_k) \tag{6}$$

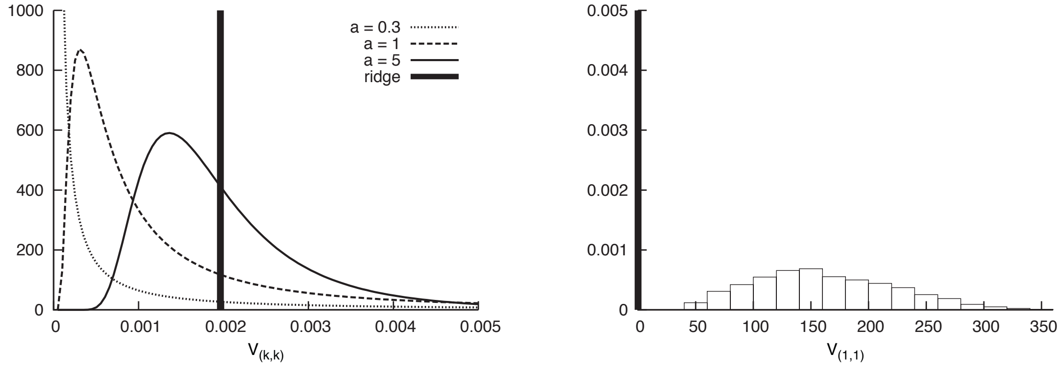
To specify the parameters a and b it is useful to consider two features of the prior: the effective model size and *shrinkage adaptivity*. Shrinkage adaptivity determines how strongly shrinkage is adapted a posteriori to each variable's performance.

Shrinkage adaptivity is related primarily to the shape parameter a. Increasing a takes probability mass away from zero and from the right tail, and shifts it towards the center of the distribution. Therefore, holding the mean of the density constant, as a increases the variance of the density decreases and thus shrinkage adaptivity also decreases.

Another way to think about shrinkage adaptivity is to note that a controls the kurtosis of the marginal prior for β, which is a Student's t-density. Low a means that the prior for β is very leptokurtic and thus puts much probability on zero and in the tails.

Given a value of a and the matrix X, the value of b can be adjusted to deliver a desired prior expected effective model size. (4) gives the effective model size for a fixed V. In the adaptive ridge model V is random and its distribution implies a distribution of the effective number of parameters. This distribution is nonstandard but can be easily simulated by Monte Carlo (cf. Hodges and Sargent, 2001).

Panel A of Figure 1 illustrates the effect of changing parameter a. It shows prior densities of a diagonal entry of V corresponding to different values of a (these densities are the same across all diagonal entries - see (6)). In all cases



A.

B.

Figure 1 – Prior distributions of diagonal entries of V implying expected effective model size of 7 in the PWT6.0 dataset. A: Prior densities of $1/\tau_k$ in the adaptive ridge model and the fixed value $1/\bar{\tau}$ in the ridge model. B: Prior distribution of $V_{(1,1)}$ in BMA. $g = 1/K^2$, approximation on the basis of 100,000 models drawn from the prior distribution of models.

b has been adjusted to produce the mean model size of 7 (the data X are taken from the baseline dataset, described later). The vertical line denotes the value of $1/\bar{\tau}$ in the ridge regression corresponding to the prior model size 7.

2.3 Bayesian Model Averaging

Bayesian model averaging (BMA) assumes dropping variables from X. The resulting regressions with subsets of the original K variables will be called *submodels*. In BMA prior probabilities are attached to submodels. These prior probabilities are updated with information about submodel fit (measured by submodel marginal likelihood) to obtain posterior probabilities of submodels. Results of interest are then computed as weighted averages across all submodels, with the weights equal to the posterior submodel probabilities.[6]

Let M_j denote submodel j ie, a regression with a subset of regressors collected in a matrix X_j. BMA uses two sets of assumptions: prior probabilities of submodels $p(M_j)$ and priors about submodel parameters $p(\alpha, \beta, \sigma^2 | M_j)$.

I use the parameters priors proposed by Fernández et al. (2001a) and applied to cross-country growth regressions by Fernández et al. (2001b). These priors are symmetric for all submodels. Priors for α and σ^2 satisfy (2) and

[6] Good references on BMA are eg, Leamer (1978) or Hoeting et al. (1999).

priors for β satisfy:

$$p(\beta^j|\alpha, \sigma^2, M_j) = N(0, \sigma^2(gX_j'X_j)^{-1}) \tag{7}$$

where β^j is the coefficient vector of X^j. The remaining entries in β are set to 0 in submodel M_j. The term $(X_j'X_j)^{-1}$ in the variance is introduced for technical reasons, to simplify computations (see Fernández et al., 2001a, p.390). g is a small positive scalar which ensures that the prior variance is large, in line with the agnostic character of the exercise. Specification of g is discussed in Fernández et al. (2001a). I take $g = 1/K^2$ as in Fernández et al. (2001b) and $g = 1/N$ as in Sala-i-Martin et al. (2004).[7]

I set prior probabilities of submodels $p(M_j)$ following Sala-i-Martin et al. (2004) (which nests the priors of Fernández et al. (2001b) as a special case). All subsets of the K variables are assigned positive probabilities, which gives rise to 2^K submodels. Each variable is included with probability p and its coefficient is set to zero with probability $1-p$. This implies that

$$p(M_j) = p^{K_j}(1-p)^{(K-K_j)}, \tag{8}$$

where K_j is the number of variables in submodel M_j. The prior expected number of variables is pK and the effective model size is[8]

$$E(J^{BMA}) = \frac{1}{1+g}pK \tag{9}$$

Another way of looking at the BMA approach is to note that it corresponds to a particular prior about matrix V in (3). According to this prior V is a random matrix which takes 2^K discrete values V_j with probabilities $p(M_j)$. Each V_j is a $K \times K$ matrix composed of zeros and entries of $(gX_j'X_j)^{-1}$ at appropriate positions. In other words, BMA amounts to using a shrinkage prior for β that is a mixture of densities.[9]

The BMA approach reflects the prior belief that the coefficient should either be zero or should be hardly shrunk at all. To see this, consider the k-th diagonal entry of V, denoted $V_{k,k}$, which is proportional to the prior variance of β_k. $V_{k,k}$ takes the value of 0 with probability $(1-p)$. The remaining probability p is distributed among the 2^{K-1} submodels that include variable

[7] The priors of Sala-i-Martin et al. (2004) are not precisely of the form (7). However, Ley and Steel (2009) show that their approach is basically equivalent to using prior (7) with $g = 1/N$.

[8] Note that combining (4) with (7) implies that the effective number of parameters in the model with K_j regressors is $\operatorname{tr} X_j((1+g)X_j'X_j)^{-1}X_j' = 1/(1+g)K_j$.

[9] Papers which explicitly formulate BMA as a mixture shrinkage prior include Geweke (1996) and Stock and Watson (2005).

k. When g is small the diagonal entries $V_{k,k}$ are large and thus correspond to a very weak shrinkage.

Figure 1 serves to compare the BMA prior for V with the ridge and adaptive ridge priors discussed earlier. Panel B shows the distribution of a diagonal entry of V when $g = 1/K^2$ and the data X is taken from the baseline dataset discussed later. For the sake of example, the distribution of the first diagonal entry is presented, but any other one compares similarly to panel A. The comparison is striking: the BMA prior of Fernández et al. (2001b) is extreme in putting all weight on either zero or on values far in the right tail of the adaptive ridge priors of the previous subsection. Note the difference of scales of panels A and B!

The comparison of the adaptive ridge model with BMA yields two observations. First, BMA is similar to an extremely adaptive ridge model. This is so, because the prior for the variance of the shrinkage prior has all its mass at zero and far in the right tail. This is similar to what happens in the adaptive ridge model in the limit, as we decrease parameter a. Second, benchmark BMA priors introduce off-diagonal terms in the variance of the prior for β, while the variance of beta in the above adaptive ridge model is diagonal. However, the off-diagonal terms enter for technical reasons only and not because of substantial prior considerations.

The computation of the posterior is easiest in the case of the ridge regression. The computational cost of a ridge regression is the same as that of an OLS regression. The adaptive ridge model requires a Monte Carlo simulation. However, the convenient gamma prior for τ_k ensures that the efficient Gibbs sampler can be used. See the Appendix for details.

BMA is computationally most challenging, because the parameter space is discrete with an enormous support. This requires somewhat more advanced tools for simulation and for convergence diagnostics (see eg, Fernández et al., 2001a, and references therein). The trick is to sample only models that have high posterior probability, and not waist time on models with negligible posterior probability. As discussed eg, in Ley and Steel (2009), the convergence of the BMA posterior simulation is quickest when g and the prior model size are both small. In this case small submodels receive most posterior weight. Then it is enough to sample small submodels, which are relatively few. However, when both the prior model size and g are large, posterior weight is spread towards larger submodels. Then more time is needed to cover the space of relevant submodels and thus convergence is slower. This puts practical limits on increasing g and prior model size.

3 Data and Model Space

The data studied in this paper are based on the dataset created by Sala-i-Martin et al. (2004), which is referred to as the SDM dataset. The SDM dataset consists of 67 variables observed for 88 countries ($K = 67, N = 88$). The dependent variable is the average growth rate of per capita GDP (gross domestic product) over the period 1960-1996. Following Ciccone and Jarociński (2010), three versions of this dataset are used. In the original SDM dataset, the initial per capita GDP (which is among the explanatory variables) and the per capita GDP growth rate are taken from Penn World Table (PWT) version 6.0. I use these original values and as alternatives I also update these two variables using data from two more recent versions of PWT (6.1 and 6.2), which reduces the number of country observations to 84 and 79 respectively. This gives rise to three different versions of the data (X, y).

I consider these three versions of the data in order to check the sensitivity of the results to different PWT data versions. Ciccone and Jarociński (2010) found, using the same data, that many results of empirical growth studies using BMA are very sensitive to the PWT version used.[10] Moreover, Johnson et al. (2009) argue that historical data in newer versions of PWT are not necessarily better than in older versions, but simply use different assumptions in the purchasing power parity adjustments. Methodological dilemmas and data availability problems are inherent in the construction of these figures. Therefore, it is important to know to what extent empirical findings are robust to the inevitable data uncertainty.

The space of models considered is indexed by two dimensions: prior effective model size and shrinkage adaptivity. Shrinkage adaptivities considered are: BMA (the most adaptive shrinkage), adaptive ridge regressions with parameter a equal to 0.3, 0.5, 1 and 5, and ridge regression (non-adaptive shrinkage). Prior effective model sizes considered are 7, 15, 20, 30 and 40. The smallest model size is 7, as in many classical empirical growth papers (Levine and Renelt, 1992; Sala-i-Martin et al., 2004). Model sizes beyond 40 seem to be impractical with the available number of observations. Fernández et al. (2001b), Ley and Steel (2009) and some other BMA papers use the prior effective model size of $K/2$. Therefore, in the case of BMA the prior effective model size is always taken to be 33.5 instead of 30, to enable direct comparison with these published results.

[10] The robustness of techniques other than BMA is studied in Hanousek et al. (2008) and Johnson et al. (2009). They find that PWT revisions affect most panel regressions, but also some cross-country regressions.

4 Fit and Robustness

4.1 Fit: Marginal Likelihoods

This subsection reports the fit of the alternative models to the cross-country growth data. I use the standard Bayesian measure of model fit that has a rigorous decision-theoretic justification: the marginal likelihood of the data. The marginal likelihood concisely summarizes the predictive performance of a model in all the out-of-sample forecasting exercises one can perform by splitting the available sample.[11] Ratios of marginal likelihoods have an interpretation of the odds that guide the optimal choice or weighting of models. Also BMA, used extensively in the empirical growth literature, is based on the marginal likelihood as a measure of submodel fit. It is therefore consistent to use marginal likelihoods also when comparing BMA with other procedures.

Table 1 reports the marginal likelihoods of all models for each of the three versions of the dataset. Three main conclusions emerge from this table.

The first conclusion is that in terms of fit, adaptive ridge and standard ridge models are attractive alternatives to the BMA specifications used in the most cited papers in the growth literature. In Table 1 BMA marginal likelihoods are lower than the best adaptive ridge marginal likelihoods in each of the three datasets.

The overall fit of the BMA procedures turns out to be quite low in spite of the fact that some of the small submodels have very good fit (marginal likelihoods and sizes of the best submodels are reported in Table A.1 in the Appendix). However, BMA procedures consider billions (2^{67}) of submodels, many of which are very poor, and the prior probability is spread over all such submodels. As a result, the overall fit of the BMA procedures turns out to be quite low. Someone obsessed just with fit might be tempted to use only the best submodel. But this would mean ignoring model uncertainty, while the whole appeal of BMA is that it accounts for model uncertainty.[12]

[11] See eg, Geweke (2005, section 2.6.2). Predictive performance is measured by the value of the predictive density at the actual data. The computation of marginal likelihoods in the present paper is explained in the Appendix.

[12] BMA marginal likelihoods could probably be improved by using coefficient priors with stronger shrinkage. The fit with $g = 1/N$ is always higher than with $g = 1/K^2$ so it may help to increase g even further. Also Eicher et al. (2009) find that cross-country growth BMA procedures with stronger shrinkage and larger prior model size have a better out-of-sample forecasting performance. The disadvantage of high g is that the computational cost of BMA becomes larger and it may even be infeasible in some cases. A second potential improvement of the BMA is to use diagonal variance in the prior for β^j, instead of the g-prior. I compared two versions of the model with all 67 variables: the ridge model and

Table 1 – Marginal likelihoods of alternative models in PWT6.0, 6.1, 6.2

prior	BMA		adaptive ridge				ridge
mod.size	g=1/K^2	g=1/N	a=0.3	a=0.5	a=1	a=5	
PWT6.0							
7	1.1E+73	1.9E+76	8.2E+76	**1.6E+77**	2.2E+76	1.4E+73	5.5E+72
15	1.6E+70	8.8E+74	2.8E+77	2.8E+78	**5.9E+78**	3.6E+77	1.4E+77
20	1.0E+68	2.3E+73	1.3E+77	4.4E+78	**2.8E+79**	1.1E+79	6.1E+78
30*	2.3E+60	-	3.9E+75	5.1E+77	2.0E+79	**9.2E+79**	**9.7E+79**
40	2.3E+55	-	7.3E+71	5.0E+74	1.4E+77	**1.1E+78**	**2.1E+78**
PWT6.1							
7	9.1E+69	**5.0E+75**	**1.0E+75**	2.5E+74	9.3E+72	5.5E+70	2.5E+70
15	2.6E+68	1.9E+75	6.1E+75	**7.2E+76**	9.9E+75	2.6E+74	1.3E+74
20	6.4E+66	9.2E+73	1.5E+76	**1.7E+77**	1.1E+77	5.4E+75	3.2E+75
30*	1.7E+60	-	1.8E+75	8.4E+76	**6.0E+77**	7.9E+76	4.3E+76
40	4.0E+55	-	5.2E+71	2.6E+74	**1.2E+76**	**1.2E+76**	4.9E+75
PWT6.2							
7	8.5E+74	**6.8E+79**	**3.2E+78**	9.0E+77	1.7E+76	2.2E+72	7.3E+71
15	1.5E+73	7.6E+78	1.8E+79	**8.5E+79**	4.6E+78	8.2E+75	2.8E+75
20	2.3E+71	2.5E+77	1.8E+79	**9.9E+79**	3.0E+79	8.1E+76	3.6E+76
30*	1.9E+64	-	6.5E+77	9.0E+78	**2.6E+79**	1.9E+77	7.5E+76
40	2.6E+59	**-	2.1E+73	9.6E+75	**1.2E+77**	8.3E+75	1.5E+75

* For BMA the prior effective model size is 33.5 instead of 30.
** Results for model sizes higher than 20 are not reported because they do not converge using the Ley and Steel (2009) software. Also Ley and Steel (2009) report convergence problems in this dataset when $g = 1/N$ and prior model size is 33.5.

The second conclusion from Table 1 is that when fitting the growth data there is a tradeoff between shrinkage adaptivity and model size. More adaptive shrinkage specifications (BMA and adaptive ridge models with low parameter a) perform better when the effective model size is small, while less adaptive shrinkage schemes perform better with larger effective model sizes. To highlight this fact, the highest marginal likelihood of the adaptive ridge and overall, in each row of the table is printed in bold font. In all three datasets the bold numbers are lined up roughly along a diagonal from the top left to the bottom right of the tables. As argued in the Introduction to this paper, to avoid omitted variables bias it is important to use possibly large model size. Table 1 suggests that when effective model size is high, less adaptive shrinkage delivers better fit.

The third conclusion is that, among the models considered in this pa-

the model with a g-prior. The ridge model strongly dominates the g-prior for all effective model sizes and all datasets. Adaptive ridge marginal likelihoods could be increased too, for example by replacing the convenient gamma prior for τ with a prior that puts less weight on extremely weak shrinkage.

per, models with 20 - 30 effective parameters have the highest marginal likelihoods. With PWT6.0, the best fitting model is the simple ridge regression with 30 effective parameters (marginal likelihood of 9.7E+79). With PWT6.1, the best fitting model is the adaptive ridge with $a = 1$ and 30 effective parameters (marginal likelihood of 6.0E+77). With PWT6.2 the best fitting model is the adaptive ridge with $a = 0.5$ and 20 effective parameters. To sum up: when general shrinkage models are used, the data favor the view that growth is a complex phenomenon affected simultaneously by many country characteristics. Only the imposition of the restriction that coefficients should either be very large or very small (as in BMA) pushes the posterior towards small models.

4.2 Robustness to Penn World Table Revisions

This subsection studies how coefficients from alternative shrinkage models differ across versions of the dataset. I generate the posterior distribution of the coefficients of all 67 variables using each model.[13] These computations are performed three times, once for each version of the dataset.

I focus on the posterior mean of the coefficients. This is the key indicator of the economic significance of a variable's relationship with growth.[14] All variables are standardized and therefore the coefficients are interpretable as the difference in the average growth rate of a country, in percentage points per annum, associated with a one standard deviation difference in the underlying variable.

Tables 2a and 2b report three statistics about changes in results across datasets: the greatest absolute change in a posterior mean across datasets

[13] The BMA results are generated with the Markov Chain Monte Carlo Model Composition sampler of Ley and Steel (2009), using software downloaded from the Journal of Applied Econometrics archive. The chain length and other settings of the sampler are left unmodified. I take $g = 1/K^2$. Correlations of visits and posterior odds are 0.98 in PWT6.0 with model size 40, 0.97 in PWT6.2 with model size 40 and well in excess of 0.99 in all remaining cases, signaling excellent convergence. The adaptive ridge results are generated with the Gibbs sampler described in the Appendix and implemented in R (R Development Core Team, 2009). 10,000 draws from the sampler are generated and every 10th draw is retained. Convergence is confirmed using the geweke.diag function from the coda package (Plummer et al., 2007). The posterior of the ridge model is the multivariate Student density provided in the Appendix for reference.

[14] The posterior mean of the regression coefficients is well defined in all considered models. Some BMA studies focus instead on the posterior probabilities of inclusion of individual variables and on the posterior means of the coefficients *conditional on their inclusion*. The posterior mean in BMA is the product of these two quantities. In ridge-type models the posterior inclusion probability is not defined.

Table 2a – Coefficient changes between PWT6.0 and PWT6.1.

prior	BMA		adaptive ridge				ridge
mod.size	g=1/K^2	g=1/N	a=0.3	a=0.5	a=1	a=5	
A. Greatest absolute change							
7	1.02	0.89	0.97	0.86	0.18	0.03	0.02
15	1.08	0.70	0.85	0.88	0.75	0.07	0.05
20	1.03	0.65	0.79	0.80	0.84	0.13	0.08
30*	0.82	0.65	0.70	0.72	0.79	0.31	0.15
40	0.75	-	0.68	0.70	0.77	0.50	0.26
B. Third greatest absolute change							
7	0.53	0.26	0.33	0.31	0.15	0.02	0.02
15	0.47	0.20	0.20	0.20	0.18	0.04	0.04
20	0.34	0.15	0.19	0.17	0.14	0.05	0.06
30*	0.26	0.21	0.17	0.16	0.15	0.08	0.09
40	0.23	-	0.21	0.19	0.21	0.16	0.14
C. Correlation							
7	0.39	0.85	0.63	0.61	0.90	0.98	0.98
15	0.55	0.93	0.84	0.80	0.74	0.97	0.97
20	0.71	0.94	0.88	0.86	0.78	0.95	0.96
30*	0.90	0.90	0.92	0.90	0.86	0.91	0.94
40	0.93	**-	0.90	0.89	0.86	0.87	0.91

*,** See the notes below Table 1. BMA coefficients for $g = 1/N$ and model size 33.5 have been similar in repeated simulations, so they are reported in spite of failed convergence diagnostics.

PWTi and PWTj ie,

$$\max_{k\in\{1\ldots K\}} \left|E(\beta_k|y^{PWTi}) - E(\beta_k|y^{PWTj})\right|,$$

the third greatest absolute change of a posterior mean across datasets and the correlation coefficient of posterior means across datasets.

The first conclusion from Tables 2a and 2b is that more adaptive shrinkage models are less robust. In the ridge and adaptive ridge model with $a = 5$ (the last two columns of the tables) coefficients changes are by far the smallest in every row of panels A and B. The same lesson emerges from panels C, which show correlations. The correlations of coefficients of the ridge and adaptive ridge model with $a = 5$ (the last two columns of the tables) are the highest in every row of panels C. Only in models of size 40 the correlations are roughly constant across shrinkage adaptivities.

The second conclusion from Tables 2a and 2b is that the disagreements

Table 2b – Coefficient changes between PWT6.0 and PWT6.2

prior	BMA		adaptive ridge				ridge
mod.size	g=1/K^2	g=1/N	a=0.3	a=0.5	a=1	a=5	
A. Greatest absolute change							
7	1.21	0.82	1.01	0.95	0.36	0.06	0.04
15	1.12	0.60	0.80	0.87	0.83	0.12	0.10
20	1.00	0.57	0.74	0.80	0.84	0.17	0.14
30*	0.72	0.66	0.68	0.71	0.81	0.31	0.22
40	0.64	-	0.70	0.73	0.81	0.54	0.31
B. Third greatest absolute change							
7	0.67	0.49	0.45	0.37	0.14	0.04	0.03
15	0.63	0.39	0.36	0.35	0.22	0.07	0.06
20	0.57	0.42	0.36	0.34	0.25	0.10	0.08
30*	0.42	0.47	0.37	0.35	0.32	0.17	0.13
40	0.39	-	0.41	0.40	0.38	0.24	0.21
C. Correlation							
7	0.19	0.62	0.49	0.51	0.79	0.92	0.93
15	0.34	0.77	0.72	0.69	0.63	0.89	0.89
20	0.47	0.80	0.77	0.75	0.68	0.85	0.87
30*	0.69	0.76	0.81	0.79	0.75	0.79	0.82
40	0.75	**-	0.78	0.77	0.74	0.72	0.76

*,** See the notes below Tables 1 and 2a.

in BMA and highly adaptive ridge are economically meaningful, while the disagreements in ridge and weakly adaptive ridge models are not. To see that these disagreements are big, note that in BMA the greatest coefficient changes are close to 1 in many cases. This means that a one standard deviation difference in some of the variables is associated with a growth difference of x% per annum in one dataset and x+1% per annum in another. A 1% disagreement about growth rates implies after 36 years (which is the span of the sample) a disagreement of more than 40% about final GDP levels. This is a substantial disagreement.

Even the third greatest changes in BMA coefficients are economically important. In small BMA models (which fit the data better than larger BMA models), the third largest coefficient change entails a 0.5% disagreement about growth rates between PWT6.0 and PWT6.2, which implies a disagreement of approximately 20% about final GDP levels.

In contrast to BMA and strongly adaptive ridge models, the absolute size of changes in less adaptive ridge models and simple ridge models is small and economically not very significant. For example, the greatest coefficient difference between PWT6.0 and PWT6.1 using ridge regression with the effective model size of 20 is 0.08. This translates into a disagreement about the GDP level after 36 years of only around 3%. Therefore, data uncertainty is a serious concern for an econometrician using very adaptive shrinkage models, such as BMA, and hardly any concern at all for an econometrician using a simple ridge model of moderate size.

5 Empirical Results From the Baseline Model

This section focuses on one baseline specification: the adaptive ridge model with effectively 30 parameters and shrinkage adaptivity parameter $a = 5$. The first subsection discusses empirical results and compares them with the BMA results. The second subsection studies the sensitivity of the baseline results to prior specification.

The choice of the baseline model is justified as follows. As argued in the introduction, it is important to control as well as possible for all variables suggested by the theory. Therefore, first, the prior effective size of the model should be rather large. Second, the prior probability of any variables dropping out is not large, since all variables are justified by growth theory which is “open-ended”. This suggests the ridge model or the adaptive ridge model with weak adaptivity.

Given these broad guidelines, the baseline model strikes a balance between fit and robustness to data revisions. Based on the robustness considerations and given the data uncertainty, only effective model sizes below 40 and adaptive ridge models with at least $a = 5$ are appealing. The specification with $a = 5$ and effective size 30 is also quite close to the best fitting models in PWT6.0 and PWT6.1.

5.1 Posterior Means of Coefficients

Table 3 reports posterior means of regression coefficients obtained with the three datasets. The first columns show the baseline model coefficients and their standard deviations. The last three columns show the BMA coefficients using the Fernández et al. (2001b) priors ie, $g = 1/K^2$ and effective model size of 33.5. This table shows 21 variables that have a coefficient of at least 0.15 in at least one case. The coefficients of the remaining variables are available from the author upon request. The first 14 rows show all variables that have a baseline model coefficient of at least 0.15 in at least one of the datasets. The subsequent 7 rows show the remaining variables that have a BMA coefficient of at least 0.15 in at least one of the datasets. The threshold of 0.15 is chosen arbitrarily, to retain only variables with an economically noticeable impact. A coefficient of 0.15 means that a two standard deviations difference in the variable in question is associated with a difference of roughly 11% in the GDP level after 36 years.

Of all the variables, the initial GDP has the strongest impact on growth across all datasets and all models. Higher initial GDP is associated with lower growth, consistently with conditional convergence. The weakest convergence is found in the PWT6.0 dataset (coefficient of -0.24 in the baseline model). The effect of initial GDP on growth is twice as strong in PWT6.1 and PWT6.2 (-0.56 and -0.53 in the baseline model). The BMA coefficients are about twice as large (-0.57, -1.35 and -1.28). Overall, in the baseline model the conditional convergence is economically significant but slower than in BMA and considerable disagreement exists across datasets.

The effect of the next six variables in the baseline model exceeds the 0.15 threshold and is consistent across datasets. Primary Schooling has a coefficient of above 0.3, which is the second largest in absolute value. Primary Schooling is also the second most important variable in BMA, where its coefficients are two to three times larger (ranging from 0.63 in PWT6.0 to 0.95 in PWT6.1).

The results for East Asian Dummy and Fraction Confucius are a good illustration of the intuitive advantage of less adaptive shrinking over BMA in presence of correlated variables. East Asian Dummy and Fraction Confucius are very similar: they are both zero in all but nine observations. A weakly adaptive ridge estimation deals with such multicollinearity by shrinking both coefficients. The posterior effect on growth gets distributed roughly equally between the two highly correlated variables. Their coefficients are stable across datasets, ranging from 0.2 to 0.3. In contrast, BMA tries to choose the better one of the two variables, which is difficult and leads to unstable results in the presence of data uncertainty. In BMA only East Asian Dummy matters in the PWT6.0 and PWT6.1 data while Fraction Confucius is irrelevant, but only Fraction Confucius matters in the PWT6.2 data while East Asian Dummy is irrelevant.

Table 3 – Posterior means of coefficients: the baseline model and the BMA of Fernández et al. (2001b).

	adaptive ridge, $a = 5$ $E(J) = 30$						BMA, $g = 1/K^2$ $E(J) = 33.5$		
	PWT60		PWT61		PWT62		PWT60	PWT61	PWT62
	mean	std	mean	std	mean	std	mean	mean	mean
GDP in 1960 (log)	-0.24	(0.18)	-0.56	(0.27)	-0.53	(0.24)	-0.57	-1.35	-1.28
Primary Schooling in 1960	0.30	(0.16)	0.30	(0.17)	0.39	(0.18)	0.63	0.95	0.92
Fraction Confucius	0.23	(0.12)	0.23	(0.12)	0.27	(0.11)	0.09	0.06	0.36
East Asian Dummy	0.29	(0.15)	0.23	(0.15)	0.20	(0.14)	0.54	0.36	0.13
Fraction Buddhist	0.19	(0.12)	0.20	(0.12)	0.22	(0.12)	0.04	0.04	0.09
Life Expectancy in 1960	0.16	(0.17)	0.16	(0.17)	0.16	(0.16)	0.23	0.18	0.01
Sub-Saharan Africa Dummy	-0.17	(0.16)	-0.22	(0.17)	-0.24	(0.17)	-0.13	-0.12	-0.55
Investment Price	-0.29	(0.12)	-0.33	(0.11)	-0.05	(0.10)	-0.36	-0.46	0.00
Fraction GDP in Mining	0.17	(0.12)	0.23	(0.13)	0.00	(0.11)	0.05	0.10	0.00
Civil Liberties	-0.14	(0.13)	-0.17	(0.14)	-0.09	(0.12)	-0.01	0.00	0.00
Openness Measure 1965-74	0.12	(0.13)	0.16	(0.14)	0.09	(0.12)	0.03	0.02	0.04
Years Open 1950-94	0.19	(0.15)	0.12	(0.14)	0.10	(0.12)	0.05	0.02	0.02
Primary Exports 1970	-0.08	(0.14)	-0.15	(0.14)	-0.13	(0.13)	-0.02	-0.10	-0.10
Real Exchange Rate Distortions	-0.16	(0.12)	-0.11	(0.12)	-0.07	(0.11)	-0.03	-0.01	0.00
Fraction of Tropical Area	-0.12	(0.15)	-0.11	(0.15)	-0.08	(0.13)	-0.39	-0.45	-0.01
Population Density Coastal in 1960s	0.08	(0.13)	0.08	(0.12)	0.00	(0.11)	0.19	0.38	0.02
Population Density 1960	0.12	(0.11)	0.14	(0.12)	0.02	(0.10)	0.03	0.29	0.00
Fertility in 1960s	-0.05	(0.17)	-0.07	(0.16)	-0.12	(0.15)	-0.01	-0.10	-0.67
Malaria Prevalence in 1960s	-0.10	(0.14)	0.02	(0.14)	-0.01	(0.13)	-0.15	0.00	0.00
Air Distance to Big Cities	-0.01	(0.13)	-0.02	(0.13)	-0.04	(0.12)	-0.01	-0.17	-0.01
Fraction Muslim	0.06	(0.12)	0.05	(0.12)	0.05	(0.11)	0.05	0.09	0.17

Fraction Buddhist is another variable that captures the rapid growth of Asian countries. Its effect on growth is positive, with a coefficient of about 0.2 consistently across datasets. This contrasts with much smaller coefficients in BMA (0.04 to 0.09).

The effect of Life Expectancy is 0.16 with all datasets. This contrasts with the BMA results, where the coefficient is 0.23 for PWT6.0, but only 0.01 for PWT6.2. The Sub-Saharan Africa Dummy is negative, with a value between -0.17 and -0.24. In contrast, in BMA the coefficient varies between -0.12 with PWT6.1 and -0.55 with PWT6.2. Overall, the coefficients of these six variables are not only among the largest in absolute value but are also very stable across datasets.

There is considerable disagreement across datasets about the next two variables, both in the baseline model and in the BMA. First, a one standard deviation increase in the relative Investment Price is associated with 0.29 and 0.33 lower growth rate according to PWT6.0 and PWT6.1, but it is irrelevant in PWT6.2. These coefficients and their disagreement are similar to the BMA results (coefficients of -0.36, -0.46 and 0). Second, a one standard deviation increase in the Fraction of GDP in the Mining sector is associated with a 0.17 and 0.23 higher growth rate according to PWT6.0 and PWT6.1, but is irrelevant according to PWT6.2. This differs from BMA where the coefficient of this variable never exceeds 0.1.

The subsequent five variables: Civil Liberties, Openness Measure 1965-74, Years Open 1950-94, Primary Exports in 1970 and Real Exchange Rate Distortions have individually small coefficients, which are around the threshold value of 0.15 and rather consistent across datasets. In contrast, the BMA coefficients of these variables are below 0.05 in all cases except Primary Exports, which has a negative coefficient of up to 0.1 in absolute value, but inconsistently across datasets.

This contrast between insignificance in BMA and moderate significance in the baseline model is even stronger when considering that three of these variables capture commitment to free trade. Openness Measure 1965-74 and Years Open 1950-94 are just different measures of openness, and Real Exchange Rate Distortions is another proxy for trade policies. These variables would usually co-move in practice, so it is interesting to consider their joint effect on growth. A country that has one standard deviation higher measures of openness and a one standard deviation lower Real Exchange Rate Distortions will grow 0.46 percentage point per annum faster according to PWT6.0, 0.39 according to PWT6.1 and 0.26 according to PWT6.2. The corresponding figures in BMA are only 0.1, 0.04 and 0.06. Thus, in contrast to BMA, the baseline model detects an economically significant association between trade openness and growth.

The BMA results for the subsequent seven variables disagree widely across datasets. The largest difference is for Fertility: its BMA coefficient is -0.67 with PWT6.2 and only -0.01 and -0.10 with PWT6.0 and PWT6.1. The effect of Fertility is also negative in the baseline model, but smaller and much more consistent (between -0.05 and -0.12). Fraction of Tropical Area has BMA coefficients of -0.39, -0.45 and only -0.01 with the three datasets. In contrast, its baseline model coefficients are again smaller and more consistent (ranging from -0.12 to -0.08). Other variables with large differences in BMA coefficients across datasets are Population Density, Population Density Coastal, Malaria Prevalence, Air Distance to Big Cities and Fraction Muslim. These and other disagreements in BMA and their sensitivity to various assumptions are studied in detail in Ciccone and Jarociński (2010). In contrast to these BMA results, the baseline model coefficients are rather small and hence fairly consistent across datasets.

Overall, Table 3 illustrates with concrete examples the advantage of the baseline model over BMA in terms of robustness to data revisions. The findings of the baseline model are also nontrivially different from the BMA findings. Controlling for more variables we obtain partial regression coefficients which are never as big in absolute value as the largest BMA coefficients. However, a number of widely discussed variables that are irrelevant in BMA have a noticeable effect in the baseline model. Proxies for open foreign trade regimes are the most notable case.

5.2 Sensitivity to Prior Specification

This section studies the sensitivity of the baseline results to prior specification. Table 4 shows correlation coefficients of baseline model posterior means with other models' posterior means. These correlations are calculated for each of the three datasets. Unsurprisingly, the correlations fall as the distance from the baseline model ($a = 5, J = 30$) increases.

Assessing the sizes of these correlations is tricky, but we have seen in Table 3 that the differences between the baseline model and the BMA of Fernández et al. (2001b) are quite substantive. The correlations in these cases are 0.75 (in PWT6.0), 0.77 (in PWT6.1) and 0.83 (in PWT6.2). Judging by this standard, all models with prior model size of 7 might differ by even more and so will many of the models with prior model size 15.

Table 5 gives a more direct idea of the economic sizes of the disagreements. It reports, for each model and each dataset, the three largest absolute differences of posterior means from the baseline model. Each entry contains three pieces of information: first, the respective model coefficient minus the baseline model coefficient; second, the acronym of the variable name (the

Table 4 – The correlation of model coefficients with the baseline model coefficients.

prior	BMA		adaptive ridge				ridge
mod.size	g=1/K^2	g=1/N	a=0.3	a=0.5	a=1	a=5	
PWT6.0							
7	0.40	0.73	0.72	0.73	0.71	0.84	0.83
15	0.56	0.78	0.87	0.90	0.91	0.93	0.92
20	0.65	0.80	0.89	0.92	0.96	0.97	0.96
30*	0.75	0.81	0.90	0.92	0.97	1.00	1.00
40	0.77	-	0.91	0.92	0.95	0.97	0.98
PWT6.1							
7	0.75	0.77	0.83	0.84	0.80	0.77	0.77
15	0.75	0.80	0.84	0.86	0.90	0.89	0.88
20	0.75	0.82	0.86	0.87	0.90	0.95	0.92
30*	0.77	0.84	0.88	0.89	0.91	1.00	0.98
40	0.79	-	0.89	0.90	0.91	0.97	0.98
PWT6.2							
7	0.81	0.82	0.85	0.87	0.86	0.69	0.71
15	0.81	0.84	0.87	0.88	0.91	0.84	0.83
20	0.82	0.84	0.88	0.89	0.91	0.92	0.89
30*	0.83	0.84	0.88	0.89	0.91	1.00	0.97
40	0.84	**-	0.88	0.89	0.90	0.96	0.97

*,** See the notes below Tables 1 and 2a.

acronyms are explained under the table) and third, the respective model's coefficient itself.

The first lesson from this table is that in the majority of cases the largest and second largest coefficient differences involve the initial GDP and Primary Schooling. The effect of the initial GDP ie, the conditional convergence tends to be stronger in larger and more adaptive models. Similarly, in larger and/or more adaptive models the effect of Primary Schooling on growth is stronger.

As regards other variables, in PWT6.0 the top three differences often involve the East Asia Dummy (in 12 cases out of 30) and Investment Price (nine cases). East Asia Dummy has a stronger positive effect in smaller and more adaptive models. Relative Investment Price has a weaker negative effect in smaller models.

Table 5 – The three largest differences of model coefficients from the baseline model coefficients.

PWT6.0

prior	BMA	adaptive ridge				ridge
mod.size	g=$1/K^2$	a=0.3	a=0.5	a=1	a=5	
7	0.64 EAST 0.92 -0.61 MALFAL66-0.70 0.26 IPRICE1 -0.03	0.35 EAST 0.64 -0.14 MALFAL66-0.24 -0.14 MINING 0.03	0.34 EAST 0.62 0.15 logy0 -0.09 -0.15 MINING 0.03	0.30 EAST 0.59 0.22 logy0 -0.02 0.17 IPRICE1 -0.12	0.24 logy0 0.00 -0.22 P60 0.08 0.21 IPRICE1 -0.09	0.24 logy0 0.00 -0.23 P60 0.07 0.22 IPRICE1 -0.08
15	0.52 EAST 0.81 -0.38 MALFAL66-0.47 -0.20 CONFUC 0.04	0.17 EAST 0.46 0.17 P60 0.47 -0.16 logy0 -0.40	0.15 EAST 0.43 0.12 P60 0.43 0.11 CIV72 -0.04	0.14 EAST 0.43 0.11 logy0 -0.13 -0.10 MINING 0.07	0.19 logy0 -0.05 -0.15 P60 0.16 0.12 IPRICE1 -0.17	0.19 logy0 -0.05 -0.17 P60 0.14 0.14 IPRICE1 -0.15
20	0.45 EAST 0.73 -0.24 MALFAL66-0.34 -0.19 CONFUC 0.06	-0.28 logy0 -0.52 0.21 P60 0.52 0.10 EAST 0.39	-0.19 logy0 -0.43 0.16 P60 0.46 0.09 EAST 0.37	0.07 GGCFD3 -0.06 0.07 CIV72 -0.08 0.07 EAST 0.36	0.15 logy0 -0.09 -0.10 P60 0.21 0.08 GGCFD3 -0.05	0.15 logy0 -0.09 -0.13 P60 0.17 0.10 IPRICE1 -0.20
30*	0.33 P60 0.64 -0.33 logy0 -0.57 -0.24 TROPICAR-0.37	-0.44 logy0 -0.68 0.24 P60 0.55 0.13 LIFE060 0.29	-0.40 logy0 -0.64 0.21 P60 0.51 0.14 LIFE060 0.30	-0.22 logy0 -0.46 0.13 P60 0.44 0.06 LIFE060 0.22	0.00 - - 0.00 - - 0.00 - -	-0.05 P60 0.26 0.04 logy0 -0.20 -0.02 EAST 0.27
40	-0.44 logy0 -0.68 0.37 P60 0.67 -0.23 TROPICAR-0.36	-0.52 logy0 -0.76 0.23 P60 0.53 0.15 LIFE060 0.31	-0.48 logy0 -0.72 0.22 P60 0.53 0.12 LIFE060 0.28	-0.36 logy0 -0.60 0.18 P60 0.49 -0.10 IPRICE1 -0.40	-0.19 logy0 -0.43 -0.10 GGCFD3 -0.23 0.10 P60 0.41	-0.12 logy0 -0.36 -0.10 GGCFD3 -0.24 -0.07 IPRICE1 -0.37

PWT6.1

prior	BMA	adaptive ridge				ridge
mod.size	g=$1/K^2$	a=0.3	a=0.5	a=1	a=5	
7	-0.49 logy0 -1.04 0.36 P60 0.68 0.32 EAST 0.54	-0.55 logy0 -1.10 0.39 P60 0.71 0.16 CIV72 -0.00	-0.39 logy0 -0.94 0.39 P60 0.71 -0.15 MINING 0.08	0.35 logy0 -0.20 0.21 EAST 0.43 -0.20 MINING 0.04	0.52 logy0 -0.03 -0.24 P60 0.08 0.22 IPRICE1 -0.10	0.53 logy0 -0.02 -0.25 P60 0.07 0.24 IPRICE1 -0.08
15	-0.68 logy0 -1.23 0.62 P60 0.94 -0.39 TROPICAR-0.50	-0.70 logy0 -1.25 0.46 P60 0.78 0.15 CIV72 -0.01	-0.62 logy0 -1.17 0.41 P60 0.73 0.15 CIV72 -0.01	-0.33 logy0 -0.88 0.25 P60 0.57 0.13 CIV72 -0.03	0.43 logy0 -0.12 -0.17 P60 0.15 -0.14 MINING 0.10	0.45 logy0 -0.10 -0.19 P60 0.13 0.16 IPRICE1 -0.16
20	-0.74 logy0 -1.29 0.67 P60 0.99 -0.39 TROPICAR-0.50	-0.76 logy0 -1.31 0.43 P60 0.75 0.14 CIV72 -0.02	-0.68 logy0 -1.23 0.40 P60 0.72 0.13 CIV72 -0.03	-0.52 logy0 -1.07 0.28 P60 0.60 0.11 CIV72 -0.05	0.33 logy0 -0.22 -0.11 P60 0.21 0.10 SAFRICA -0.14	0.39 logy0 -0.16 -0.15 P60 0.17 0.11 IPRICE1 -0.21
30*	-0.84 logy0 -1.39 0.68 P60 1.00 -0.36 TROPICAR-0.46	-0.83 logy0 -1.38 0.42 P60 0.74 -0.13 SAFRICA -0.37	-0.81 logy0 -1.36 0.39 P60 0.71 -0.11 SAFRICA -0.34	-0.70 logy0 -1.25 0.31 P60 0.63 -0.08 SAFRICA -0.31	0.00 - - 0.00 - - 0.00 - -	0.20 logy0 -0.35 -0.06 P60 0.26 0.03 SAFRICA -0.21
40	-0.88 logy0 -1.43 0.67 P60 0.98 -0.30 TROPICAR-0.40	-0.89 logy0 -1.44 0.42 P60 0.74 -0.23 SAFRICA -0.46	-0.87 logy0 -1.42 0.38 P60 0.70 -0.23 SAFRICA -0.46	-0.82 logy0 -1.37 0.33 P60 0.65 -0.21 SAFRICA -0.44	-0.38 logy0 -0.93 -0.14 SAFRICA -0.37 0.13 P60 0.45	0.13 MALFAL66 0.14 -0.11 CIV72 -0.27 -0.11 GGCFD3 -0.22

Table 5 – - Continued.

PWT6.2

prior	BMA	adaptive ridge				ridge
mod.size	g=1/K^2	a=0.3	a=0.5	a=1	a=5	
7	-0.68 logy0 -1.23 -0.52 FERTLDC1-0.63 0.38 P60 0.79	-0.59 logy0 -1.14 -0.40 FERTLDC1-0.51 0.35 P60 0.76	-0.49 logy0 -1.04 0.32 P60 0.73 -0.26 FERTLDC1-0.37	0.27 EAST 0.47 0.18 logy0 -0.38 0.12 GOVNOM1-0.02	0.53 logy0 -0.02 -0.34 P60 0.07 0.19 SAFRICA -0.06	0.53 logy0 -0.02 -0.34 P60 0.06 0.19 SAFRICA -0.06
15	-0.71 logy0 -1.26 -0.55 FERTLDC1-0.67 0.44 P60 0.84	-0.65 logy0 -1.20 0.43 P60 0.83 -0.34 FERTLDC1-0.46	-0.61 logy0 -1.17 0.41 P60 0.82 -0.26 FERTLDC1-0.38	-0.41 logy0 -0.97 0.33 P60 0.74 -0.09 FERTLDC1-0.20	0.45 logy0 -0.10 -0.26 P60 0.14 0.13 SAFRICA -0.11	0.47 logy0 -0.09 -0.28 P60 0.13 0.14 SAFRICA -0.11
20	-0.71 logy0 -1.26 -0.55 FERTLDC1-0.67 0.46 P60 0.86	-0.70 logy0 -1.25 0.47 P60 0.88 -0.34 FERTLDC1-0.46	-0.67 logy0 -1.23 0.45 P60 0.86 -0.28 FERTLDC1-0.39	-0.52 logy0 -1.07 0.38 P60 0.78 -0.11 FERTLDC1-0.23	0.36 logy0 -0.19 -0.19 P60 0.22 0.10 SAFRICA -0.14	0.41 logy0 -0.15 -0.23 P60 0.17 0.11 SAFRICA -0.14
30*	-0.73 logy0 -1.28 -0.56 FERTLDC1-0.67 0.51 P60 0.92	-0.81 logy0 -1.37 0.54 P60 0.94 -0.33 FERTLDC1-0.44	-0.80 logy0 -1.35 0.51 P60 0.92 -0.27 FERTLDC1-0.38	-0.71 logy0 -1.27 0.45 P60 0.86 -0.16 FERTLDC1-0.28	0.00 - - 0.00 - - 0.00 - -	0.23 logy0 -0.32 -0.12 P60 0.29 0.04 SAFRICA -0.21
40	-0.76 logy0 -1.31 0.55 P60 0.96 -0.55 FERTLDC1-0.66	-0.91 logy0 -1.47 0.60 P60 1.01 -0.34 FERTLDC1-0.45	-0.90 logy0 -1.46 0.59 P60 1.00 -0.28 FERTLDC1-0.40	-0.86 logy0 -1.41 0.54 P60 0.95 -0.21 SAFRICA -0.46	-0.42 logy0 -0.97 0.22 P60 0.63 -0.14 SAFRICA -0.39	-0.11 DENS65C -0.10 -0.10 SAFRICA -0.35 -0.09 NEWSTATE-0.15

* For BMA the prior effective model size is 33.5 instead of 30.

Acronyms: EAST - East Asia Dummy, logy0 - GDP in 1960 (log), MALFAL66 - Malaria Prevalence in 1960's, P60 - Primary Schooling in 1960, IPRICE1 - Investment Price, MINING - Fraction GDP in Mining, CONFUC - Fraction Confucius, CIV72 - Civil Liberties (Gastil index), GGCFD3 - Public Investment Share, TROPICAR - Fraction of Tropical Area, LIFE60 - Life Expectancy in 1960, SAFRICA - Sub-Saharan Africa Dummy, FERTLDC1 - Fertility in 1960, GOVNOM1 - Nominal Government GDP Share 1960s, DENS65C - Population Density Coastal in 1960s, NEWSTATE - Timing of Independence.

In PWT6.1 the top three differences often involve Sub-Saharan Africa Dummy (nine cases) and Civil Liberties (eight cases). Sub-Saharan Africa Dummy has a stronger negative effect in large models. Civil Liberties have a weaker negative effect in more adaptive models.

In both PWT6.0 and 6.1 Tropical Area has a stronger negative effect in more adaptive models, while the Fraction of GDP in Mining has a weaker positive effect in smaller models.

In PWT6.2 the top three differences often involve Fertility (18 cases) and again Sub-Saharan Africa Dummy (ten cases). Fertility has a very strong negative effect in more adaptive shrinkage models, while it has no such effect in the baseline. Sub-Saharan Africa Dummy has a weaker effect in smaller models.

The overall conclusion from Table 5 is that effective model size and shrinkage adaptivity matter a lot for the results. The lessons from more adaptive and smaller models are often quite different regarding the most important variables. However, researchers who agree that model sizes should be at least 20 and shrinkage adaptivity a at least one will find reasonably similar results, except for a slight disagreement about convergence and Primary Schooling.

The simple ridge model is the closest one to the baseline model. Researchers using the ridge model would draw virtually the same conclusions from the data, except for a slightly lower convergence speed.

6 Conclusions

This paper proposes a new approach to the empirical growth research. In contrast to much of the literature, this paper does not view this research as a competition of small models. Instead, it models growth as a product of many mutually canceling or reinforcing factors. The Introduction argues that this view is consistent with the existing growth theory. The rest of the paper shows that this view is also empirically plausible. Moreover, it delivers results robust to the measurement error inherent in the data.

The good news is that the approach of this paper is computationally very simple. This paper shows that a robust analysis of a large cross-country dataset can be performed with a simple ridge regression, which is available in most econometric packages and involves only one matrix inversion. This allows empirical growth researchers to shift their attention from computational issues to the other challenges facing empirical growth research, such as endogeneity of growth determinants, nonlinearity and new data collection.

Appendix: Computation details

A.1 The Posterior of the Ridge Model

The kernel of the posterior of the ridge model is given by the standard textbook formula, provided here for completeness:

$$p(\beta|y) \propto \left((\beta - \bar{\beta})'(X'X + \mathrm{diag}(\tau))(\beta - \bar{\beta}) + s\right)^{-\frac{N+K-3}{2}} \tag{A.1}$$

where $\bar{\beta} = (X'X + \mathrm{diag}(\tau))^{-1}X'y$ and $s = y'y - y'X(X'X + \mathrm{diag}(\tau))^{-1}X'y - N\bar{y}^2$. (A.1) is a kernel of the multivariate Student density with mean $\bar{\beta}$ and variance:

$$\mathrm{Var}(\beta) = \frac{s}{N-5}(X'X + \mathrm{diag}(\tau))^{-1}$$

A.2 Gibbs Sampler for the Adaptive Ridge Model

The joint posterior of all parameters is proportional to the product of the kernels of the likelihood, the Normal prior for β, the gamma prior for τ_k and the noninformative priors for σ^2 and α. Since α is not of interest, I integrate it out analytically. The posterior kernel of the remaining parameters is:

$$\begin{aligned} p(\beta, \sigma^2, \tau_1 \ldots \tau_K) \propto & (\sigma^2)^{-(N-1)/2} \exp\left(-\frac{1}{2}\frac{(\tilde{y} - X\beta)'(\tilde{y} - X\beta)}{\sigma^2}\right) \\ & \times \prod_{k=1}^{K} (\sigma^2)^{-\frac{1}{2}} \tau_k^{-1/2} \exp\left(-\frac{1}{2}\frac{\beta_k^2 \tau_k}{\sigma^2}\right) \times \prod_{k=1}^{K} \tau_k^{a-1} \exp\left(-b\tau_k\right) \times \frac{1}{\sigma^2} \end{aligned} \tag{A.2}$$

where $\tilde{y}$ is demeaned y. I assume throughout the paper that X has been demeaned already. Denoting the set of all parameters to be estimated as $\Theta \equiv \{\beta, \tau_1, \ldots, \tau_K, \sigma^2\}$ the conditional posteriors are as follows:

$$\begin{aligned} p(\beta|y, X, \Theta\backslash\{\beta\}) & \propto \exp\left(-\frac{1}{2}\left(\beta'\frac{X'X}{\sigma^2}\beta - 2\frac{\beta'X'\tilde{y}}{\sigma^2} + \beta'\frac{\mathrm{diag}(\tau_1 \ldots \tau_K)}{\sigma^2}\beta\right)\right) \\ & \propto \mathrm{N}\left((X'X + \mathrm{diag}(\tau_1 \ldots \tau_K))^{-1}X'\tilde{y}, \sigma^2(X'X + \mathrm{diag}(\tau_1 \ldots \tau_K \right. \end{aligned} \tag{A.3}$$

$$p(\tau_k|y, X, \Theta\backslash\{\tau_k\}) \propto \tau_k^{a+\frac{1}{2}-1} \exp\left(-\left(b + \frac{\beta_k^2}{2\sigma^2}\right)\tau_k\right) \propto \mathrm{G}\left(a + \frac{1}{2}, b + \frac{\beta_k^2}{2\sigma^2}\right) \tag{A.4}$$

$$p(\sigma^2|y, X, \Theta\backslash\{\sigma^2\}) \propto (\sigma^2)^{-(N+K-1)/2} \exp\left(-\frac{1}{2}\frac{(y-X\beta)'(y-X\beta)+\beta'\operatorname{diag}(\tau_1\ldots\tau_K)\beta}{\sigma^2}\right)$$
$$\propto \mathrm{IG}_2\left((y-X\beta)'(y-X\beta)+\beta'\operatorname{diag}(\tau_1\ldots\tau_K)\beta, N+K-3\right) \qquad \text{(A.5)}$$

A sample from the posterior is easily generated with the Gibbs sampler ie, by repeatedly drawing in turn from (A.3), (A.4) and (A.5).

A.3 Computation of Marginal Likelihoods

This subsection explains the computation of the marginal likelihoods. Special care is taken in the computation to make sure that marginal likelihoods are comparable across models. Because of the improper priors in (2), the levels of marginal likelihoods are not interpretable. Marginal likelihoods are only determined up to an arbitrary multiplicative factor coming from the improper part of the prior. However, the improper part of the prior is common to all models considered in this paper and therefore it does not affect comparisons across models. Second, because of the common structure of all models given in (1), (2) and (3), there are other common factors which can be omitted. The computations below ensure that any omitted multiplicative factors are the same. Therefore the ratios of these marginal likelihoods (Bayes factors) are meaningful and allow model comparisons. This is the same as in the BMA of Fernández et al. (2001a) or Sala-i-Martin et al. (2004): the levels of marginal likelihoods are not interpretable, but their relative sizes are meaningful and determine model weights.

Models satisfying (1), (2) and (3) differ only in the V matrix. Therefore, the factor of the marginal likelihoods that does not cancel in the odds ratio for any pair of such models is:

$$p(y) \propto |X'X+V^{-1}|^{-1/2}|V|^{-1/2}\left(y'y - y'X(X'X+V^{-1})^{-1}X'y - N\bar{y}^2\right)^{-(N-1)/2} \qquad \text{(A.6)}$$

This expression is obtained by computing

$$p(y|V) = \int p(y|\alpha,\beta,\sigma^2,X)dp(\alpha)p(\beta)p(\sigma^2)$$

and dropping all terms that do not depend on V and therefore cancel in an odds ratio. See Fernández et al. (2001a) for a similar expression.

Ridge regression. The marginal likelihood of a ridge regression is computed directly by evaluating (A.6).

When V is stochastic then in order to compute the marginal likelihood we need to integrate it out using its prior distribution.

Adaptive ridge. In the adaptive ridge regression we have $V = \text{diag}(\tau)^{-1}$ where τ is a vector of gamma random variables. I integrate out τ with Monte Carlo ie, repeatedly drawing τ from its prior distribution and averaging the values of (A.6) across draws. This procedure converges very quickly when parameter a is high, but for $a < 1$ many draws are needed. The results in Table 1 are obtained with 10 million draws. The reason is that when a is low, the distribution of τ is more spread out, so that it covers models with very different fit and it takes more time to explore it.

BMA. In principle, one could use the same Monte Carlo computation as before. However, in BMA the distribution of V is discrete over 2^K points. When K is large this Monte Carlo computation would converge too slowly. Therefore, the weight of the BMA procedure as a whole was computed in a different way, utilizing the output of the BMA software provided by Ley and Steel (2009).[15]

Let us define some notation first. As discussed in section 2.3, a BMA procedure depends on a number of specifications: priors about parameters given submodels and prior probability of submodels. Let $\mathcal{B}$ denote a particular specification of all these assumptions. Let M_1 denote the submodel with the highest marginal likelihood. The posterior probability of submodel M_1 conditional on the BMA procedure $\mathcal{B}$ satisfies:

$$p(M_1|y,\mathcal{B}) = \frac{p(M_1|\mathcal{B})p(y|M_1,\mathcal{B})}{\sum_{j=1}^{2^K} p(M_j|\mathcal{B})p(y|M_j,\mathcal{B})} = \frac{p(M_1|\mathcal{B})p(y|M_1,\mathcal{B})}{p(y|\mathcal{B})}$$

where $p(M_j|\mathcal{B})$ is the prior probability of submodel M_j in the BMA procedure $\mathcal{B}$ and $p(y|M_j,\mathcal{B})$ is the marginal likelihood of submodel M_j. This implies that

$$p(y|\mathcal{B}) = \frac{p(M_1|\mathcal{B})p(y|M_1,\mathcal{B})}{p(M_1|y,\mathcal{B})}. \tag{A.7}$$

I compute the marginal likelihood of BMA using (A.7). I take the posterior probability of the best submodel $p(M_1|y,\mathcal{B})$ from the output of the BMA software of Ley and Steel (2009). This software reports also which variables enter the best submodel M_1. Knowing the composition of M_1 the two remaining quantities are easy to compute. I compute the marginal likelihood $p(y|M_1,\mathcal{B})$ using (A.6), replacing X with X_j and V with $(gX_j'X_j)^{-1}$. I compute the prior probability $p(M_1|\mathcal{B})$ using (8). Table A.1 reports all the quantities used to compute the entries of Table A.6.

[15] This software uses advanced Monte Carlo methods to explore only the most relevant part of the space of 2^K models - see their paper and the literature quoted therein.

Table A.1 – BMA best submodels (M_1): marginal likelihood $p(y|M_1, \mathcal{B})$, posterior probability $p(M_1|y, \mathcal{B})$ and number of regressors K_1

prior	$g = 1/K^2$			$g = 1/N$		
mod.size	$p(y\mid M_1, \mathcal{B})$	$p(M_1\mid y, \mathcal{B})$	K_1	$p(y\mid M_1, \mathcal{B})$	$p(M_1\mid y, \mathcal{B})$	K_1
PWT6.0						
7	4.8E+77	0.3772	2	7.9E+83	0.0664	6
15	4.8E+77	0.1088	2	7.9E+83	0.0219	6
20	2.3E+79	0.0644	6	7.9E+83	0.0099	6
33.5	2.3E+79	0.0657	6	-		
40	2.3E+79	0.0373	6	-		
PWT6.1						
7	6.0E+77	0.1031	6	1.7E+85	0.0725	8
15	1.5E+79	0.1161	7	7.8E+85	0.0246	9
20	1.5E+79	0.1215	7	7.8E+85	0.0188	9
33.5	1.5E+79	0.0599	7	-		
40	1.2E+79	0.0369	7	-		
PWT6.2						
7	8.2E+82	0.1503	6	2.0E+87	0.0451	6
15	8.2E+82	0.1338	6	2.0E+87	0.0063	6
20	8.2E+82	0.1015	6	2.5E+89	0.0040	11
33.5	8.2E+82	0.0297	6	-		
40	8.2E+82	0.0118	6	**-		

Notes: $p(M_1|y, \mathcal{B})$ and K_1 are taken from the output of the BMA software of Ley and Steel (2009). $p(y|M_1, \mathcal{B})$ is computed using (A.6). ** See the notes below Table 1.

References

Brock, W. A. and Durlauf, S. N. (2001). Growth empirics and reality. *World Bank Economic Review*, 15(2):229–272.

Ciccone, A. and Jarociński, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics*. forthcoming.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

Denison, D. G. T. and George, E. I. (2001). Bayesian prediction using adaptive ridge estimators. Technical Report, Dept of Statistics, The Wharton School, PA.

Durlauf, S. N., Johnson, P. A., and Temple, J. R. (2005). Growth econometrics. volume 1 of *Handbook of Economic Growth*, chapter 8, pages 555–677. Elsevier.

Eicher, T. S., Papageorgiou, C., and Raftery, A. (2009). Default priors and predictive performance in bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*. forthcoming.

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

Fernández, C., Ley, E., and Steel, M. F. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5):563–76.

Gelman, A. B., Carlin, J. S., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall Texts in Statistical Science. Chapman and Hall/CRC, second edition.

Geweke, J. (1996). Variable selection and model comparison in regression. *Bayesian Statistics*, 4.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Hoboken, New Jersey, first edition.

Hanousek, J., Hajkova, D., and Filer, R. K. (2008). A rise by any other name? sensitivity of growth regressions to data source. *Journal of Macroeconomics*, 30(3):1188–1206.

Hauk, W. and Wacziarg, R. (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth*, 14(2):103–147.

Hendry, D. F. and Krolzig, H.-M. (2004). We ran one regression. *Oxford Bulletin of Economics and Statistics*, 66:799–810.

Hodges, J. S. and Sargent, D. J. (2001). Counting degrees of freedom in hierarchical and other richly-parametrized models. *Biometrica*, 88(2):367–379.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417.

Johnson, S., Larson, W., Papageorgiou, C., and Subramanian, A. (2009). At your own risk: Health warnings for growth data. paper presented at the 2009 North American Summer Meeting of Econometric Society in Boston University.

Leamer, E. E. (1978). *Specification Searches.* New York: John Wiley and Sons.

Levine, R. E. and Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, 82(4):942–63.

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(4):651–674.

Magnus, J., Powell, O., and Prufer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics*, 154:139–153.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2007). *coda: Output analysis and diagnostics for MCMC.* R package version 0.13-1.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *The American Economic Review*, 94(4):813–835.

Sims, C. A. (2003). Probability models for monetary policy decisions. Manuscript, Princeton University, available at http://sims.princeton.edu/yftp/Ottawa/ProbModels.pdf.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B*, 64(4):583–639.

Stock, J. H. and Watson, M. W. (2005). An empirical comparison of methods for forecasting using many predictors. manuscript.

Strawderman, W. E. (1978). Minimax adaptive generalized ridge regression estimators. *Journal of the American Statistical Association*, 73(363):623–627.

Wagner, M. and Hlouskova, J. (2009). Growth regressions, principal components and frequentist model averaging. Economics Series 236, Institute for Advanced Studies.