WP/99/101

INTERNATIONAL MONETARY FUND

Fiscal Affairs Department

We Just Averaged over Two Trillion Cross-Country Growth Regressions

Prepared by Eduardo Ley and Mark F.J. Steel *

Authorized for distribution by Liam P. Ebrill

July 1999

IMF WORKING PAPER

INTERNATIONAL MONETARY FUND

IMF Working Paper

WP/99/101

INTERNATIONAL MONETARY FUND

Fiscal Affairs Department

We Just Averaged over Two Trillion Cross-Country Growth Regressions

Prepared by Eduardo Ley and Mark F.J. Steel *

Authorized for distribution by Liam P. Ebrill

July 1999

Abstract

We investigate the issue of model uncertainty in cross-country growth regressions using Bayesian model averaging (BMA). We find that the posterior probability is distributed among many models, suggesting the superiority of BMA over any single model. Out-of-sample predictive results support that claim. In contrast with Levine and Renelt (1992), our results broadly support the more "optimistic" conclusion of Sala-i-Martin (1997b), namely, that some variables are important regressors for explaining cross-country growth patterns. However, the variables we identify as most useful for growth regression differ substantially from Sala-i-Martin's results.

JEL Classification Numbers: O49, C11, C52

Keywords: Bayesian model averaging, choice of regressors, economic growth, Markov chain Monte Carlo, prediction

Authors' E-Mail Addresses: eley@imf.org; Mark.Steel@ed.ac.uk

---

* Eduardo Ley is an Economist, Tax Policy Division, Fiscal Affairs Department, IMF, and Mark Steel is Professor of Economics, University of Edinburgh, Scotland, United Kingdom.

Contents Page

I. INTRODUCTION

Many empirical studies of the growth of countries attempt to identify the factors explaining the differences in growth rates by regressing observed GDP growth on a host of country characteristics that could possibly affect growth. This line of research was heavily influenced by Kormendi and Meguire (1985) and Barro (1991). Excellent recent surveys of these cross-section studies and their role in the broader context of economic growth theory are provided in Durlauf and Quah (1999) and Temple (1999).

In this paper we focus on cross-country growth regressions and attempt to shed further light on the importance of such models for empirical growth research. Prompted by the proliferation of possible explanatory variables in such regressions and the relative absence of guidance from economic theory as to which variables to include, Levine and Renelt (1992) investigate the "robustness" of the results from such linear regression models. They use a variant of the Extreme-Bounds Analysis (EBA) introduced in Leamer (1983, 1985) and conclude that very few regressors pass the extreme-bounds test. In response to this rather negative finding, Sala-i-Martin (1997b) employs a less severe test for the importance of explanatory variables in growth regressions. He considers the distance of the point estimates from zero, averaged over a set of regression models. Broadly speaking, if the averaged[1] 90 percent confidence interval of a regression coefficient does not include zero, Sala-i-Martin will classify the corresponding regressor as a variable that is strongly correlated with growth. On the basis of this methodology, Sala-i-Martin (1997b) identifies a relatively large number of variables as important for growth regression.

Here we set out to investigate this issue in a formal statistical framework that explicitly allows for the specification uncertainty described above. In particular, a Bayesian framework allows us to deal with both model and parameter uncertainty in a straightforward and formal way. We also consider an extremely large set of possible models by allowing for any subset of up to 41 regressors to be included in the model. This means we have a set of $2^{41} = 2.2 \times 10^{12}$ (over two trillion!) different models to deal with. Novel Markov chain Monte Carlo (MCMC) techniques are adopted to solve this numerical problem, using the so-called Markov chain Monte Carlo Model Composition (MC^3) technology, first used in Madigan and York (1995).

Our findings are based on the same data as those of Sala-i-Martin[2] and broadly support the more "optimistic" conclusion of Sala-i-Martin (1997b), namely that

[1] In one case he considers the confidence interval on the basis of the averaged estimates, and in another case he averages over confidence intervals. Averaging over models is either done uniformly, or with weights proportional to the likelihoods. See also footnote 11 in this context.

[2] We thank Xavier Sala-i-Martin for making the data publicly available through his website at http: //www.columbia.edu/~xs23/data/millions.htm.

some variables are important regressors for explaining cross-country growth patterns. However, the variables we identify as most useful for growth regression differ substantially from his results. In addition, we do not advocate selecting a subset of the regressors, but we use Bayesian Model Averaging (BMA), where all inference is averaged over models, using the corresponding posterior model probabilities. It is important to point out that our methodology allows us to go substantially further than the previous studies, in that we provide a clear interpretation of our results and a formal statistical basis for inference, both posterior and predictive. Finally, let us note that this paper is solely intended to investigate the issue of choice of regressors and inference in cross-country growth regressions. It does not attempt to address here the myriad of other interesting topics, such as convergence of countries, data quality *etc.*

II. THE MODEL AND THE METHODOLOGY

Following the analyzes in Levine and Renelt (1992) and Sala-i-Martin (1997b) as well as the tradition in the growth regression literature, we will consider linear regression models where GDP growth for n countries, grouped in a vector y, is regressed on an intercept, say α, and a number of explanatory variables chosen from a set of k variables in a matrix Z of dimension $n \times k$. Throughout, we assume that $\text{rank}(\iota_n : Z) = k + 1$, where ι_n is an n-dimensional vector of 1's.

Whereas Levine and Renelt and Sala-i-Martin restrict the set of regressors to always contain certain key variables and then allow for four[3] other variables to be added, we shall allow for any subset of the variables in Z to appear in the model. This results in 2^k possible models, which will thus be characterized by the selection of regressors. We call model M_j the model with the $0 \leq k_j \leq k$ regressors grouped in Z_j, leading to

$$y = \alpha\iota_n + Z_j\beta_j + \sigma\varepsilon, \tag{1}$$

where $\beta_j \in \Re^{k_j}$ groups the relevant regression coefficients and $\sigma \in \Re_+$ is a scale parameter. Furthermore, we shall assume that ε follows an n-dimensional Normal distribution with zero mean and identity covariance matrix.

In our Bayesian framework, we need to complete the above sampling model with a prior distribution for the parameters in M_j, namely α, β_j and σ. Fernández, Ley, and Steel (1998) study the issue of choosing the prior distribution in such a way that the prior is little informative on the resulting Bayes factors. They propose to use improper noninformative priors for the parameters that are common to all models, namely α and σ and a g-prior structure for β_j:

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

[3] Levine and Renelt (1992) consider one up to four added regressors, Sala-i-Martin (1997b) restricts the analysis to exactly four extra regressors.

where $f_N^q(w \mid m, V)$ denotes the density function of a q-dimensional Normal distribution on w with mean m and covariance matrix V. Fernández, Ley, and Steel (1998) conclude on the basis of a simulation study and theoretical properties that a reasonable choice for g_{0j} is given by $g_{0j} = \frac{1}{n}$, for cases where the number of observations, n, is relatively small (like in the present application).

In order to fully specify the prior distribution under model M_j, we group the irrelevant components of β under M_j in a vector $\beta_{\sim j} \in \Re^{k-k_j}$. The latter vector follows a Dirac distribution at zero, *i.e.*,

$$P_{\beta_{\sim j} \mid \alpha, \beta_j, \sigma, M_j} = P_{\beta_{\sim j} \mid M_j} = \text{ Dirac at } (0, \ldots, 0). \tag{3}$$

Denoting the space of all 2^k possible models by $\mathcal{M}$, we finally put a prior distribution over this model space $\mathcal{M}$

$$P(M_j) = p_j, \quad j = 1, \ldots, 2^k, \quad \text{with } p_j > 0 \text{ and } \sum_{j=1}^{2^k} p_j = 1. \tag{4}$$

We can now use the Bayesian paradigm to deal with model uncertainty in a perfectly straightforward way, since the posterior or predictive distribution of any quantity of interest, say Δ, is a mixture of the posterior or predictive distributions of that quantity under each of the models with mixing probabilities given by the posterior model probabilities. Thus

$$P_{\Delta \mid y} = \sum_{j=1}^{2^k} P_{\Delta \mid y, M_j} P(M_j \mid y). \tag{5}$$

This procedure, which is typically referred to as Bayesian model averaging (BMA), immediately follows from the rules of probability theory—see, *e.g.*, Leamer (1978).

The posterior model probabilities used in (5) are given by

$$P(M_j \mid y) = \frac{l_y(M_j) P(M_j)}{\sum_{h=1}^{2^k} l_y(M_h) P(M_h)}, \tag{6}$$

where $l_y(M_j)$, the marginal likelihood of model M_j, is obtained as

$$l_y(M_j) = \int p(y \mid \alpha, \beta_j, \sigma, M_j) p(\alpha, \beta_j, \sigma \mid M_j) \, d\alpha \, d\beta_j \, d\sigma, \tag{7}$$

with $p(y \mid \alpha, \beta_j, \sigma, M_j)$ the sampling model in (1) and $p(\alpha, \beta_j, \sigma \mid M_j)$ the prior defined in (2). Fernández, Ley, and Steel (1998) show that for the particular Bayesian

model in (1) – (4) the ratio of the marginal likelihoods, usually known as the Bayes factor, $B_{js} = l_y(M_j)/l_y(M_s)$, can be computed analytically. In the somewhat simplifying case where we assume, without loss of generality, that the regressors are transformed such that $\iota_n' Z = 0$, and defining $X_j = (\iota_n : Z_j)$, $\overline{y} = \iota_n' y/n$ and $M_{X_j} = I_n - X_j(X_j'X_j)^{-1}X_j'$, they obtain

$$B_{js} = \left(\frac{g_{0j}}{g_{0j}+1}\right)^{k_j/2} \left(\frac{g_{0s}+1}{g_{0s}}\right)^{k_s/2} \times \left(\frac{\frac{1}{g_{0s}+1} y' M_{X_s} y + \frac{g_{0s}}{g_{0s}+1}(y-\overline{y}\iota_n)'(y-\overline{y}\iota_n)}{\frac{1}{g_{0j}+1} y' M_{X_j} y + \frac{g_{0j}}{g_{0j}+1}(y-\overline{y}\iota_n)'(y-\overline{y}\iota_n)}\right)^{(n-1)/2}, \tag{8}$$

if $k_j \geq 1$ and $k_s \geq 1$. If one of the models, say M_j, is the model with only the intercept (*i.e.*, $k_j = 0$), the Bayes factor is simply obtained as the limit of B_{js} in (8) as g_{0j} tends to infinity.

In practice, computing the relevant posterior or predictive distribution through (5) and (6) is hampered by the very large amount of terms in the sums. In the most demanding case (subsection 3.2), we have $k = 41$ possible regressors, and we would thus need to calculate posterior probabilities for each of the $2^{41} = 2.2 \times 10^{12}$ models and average the required distributions over all these models. In order to substantially reduce this prohibitive computational effort, we shall approximate the posterior distribution on the model space $\mathcal{M}$ by simulating a sample from it, applying the MC^3 methodology of Madigan and York (1995). This consists in a Metropolis-Hastings algorithm—see, *e.g.*, Chib and Greenberg (1996)—to generate drawings from a Markov chain on $\mathcal{M}$ which has the posterior model distribution as its stationary distribution. Raftery, Madigan, and Hoeting (1997) and Fernández, Ley, and Steel (1998) use MC^3 methods in the context of the linear regression model.

In the implementation, we shall take advantage of the fact that for each pair of models, the Bayes factor can be computed analytically through (8). Thus, we shall use the chain to merely indicate which models should be taken into account in computing the sums in (5) and (6). This idea was introduced as Bayesian Random Search (BARS) in Lee (1996). In addition, Fernández, Ley, and Steel (1998) propose to use this as a convenient indicator for convergence of the chain. A high positive correlation between the posterior model probabilities based on the empirical frequencies of visits on the one hand, and the exact formula in (8) on the other hand, provides strong evidence of convergence.

III. Posterior Results

We take the same data as used and described in Sala-i-Martin (1997b), covering 140 countries, for which average per capita GDP growth was computed over the period 1960–92. Sala-i-Martin starts with the model in (1) and a large set of 62 variables that could serve as regressors. He then restricts his analysis to those models where three specific variables are always included (these are the level of GDP, life expectancy, and primary school enrollment rate, all for 1960) and for each of the remaining 59 variables he considers the estimated regression coefficient, while adding all possible triplets of the remaining 58 variables. He finally computes CDF(0) for this coefficient, which reflects the distance from zero (in the predominant direction) of the estimated values, averaged over all these possible regressions, to conclude that 22 of the 59 variables are "significant," in that CDF(0) is at least 0.95. Thus, he considers 455,126 different models,[4] which we will denote by $\mathcal{M}_S$.

A. The Reduced Set of Regressors

We shall first undertake our analysis on the basis of the 25 variables that Sala-i-Martin (1997b) flagged as being important (his three retained variables and the 22 variables in his Table 1, page 181). We have available $n = 72$ observations for all these regressors. Thus, Z will be the 72×25 design matrix corresponding to these variables (transformed by subtracting the mean, so that $\iota_n' Z = 0$), and we shall allow for any subset of these 25 regressors, leading to a total set of $2^{25} = 33.6$ million models under consideration in $\mathcal{M}$. Since we do not start from the full set of 62 variables, we do not cover all models in $\mathcal{M}_S$. However, we cover the subset of $\mathcal{M}_S$ that corresponds to the regressors favored by Sala-i-Martin. This intersection between $\mathcal{M}$ and $\mathcal{M}_S$ consists of 7,315 models and will be denoted by $\mathcal{M}_I$ in the sequel. In view of the results in Sala-i-Martin (1997b), we would certainly expect that a relatively large fraction of the posterior mass in $\mathcal{M}_S$ is concentrated in $\mathcal{M}_I$.

To analyze these data, we use the Bayesian model in (1)–(4) with a Uniform prior on model probabilities, *i.e.*, $p_j = 2^{-k}$ in (4). We run the MC^3 sampler for 100,000 draws as a burn-in and record the next 500,000 draws in model space $\mathcal{M}$. That the sampler has converged is confirmed by the very high correlation between the frequencies of recorded model visits and the probabilities based on (8) (it equals 0.975 in a representative run). In addition, various long runs from randomly chosen starting values led to virtually identical results. In a typical run, we visit 67,684 models (*i.e.*, about 0.2 percent of the total number) and posterior model probability is quite spread: the best model, which contains 13 regressors, obtains a posterior

[4] Note that the (almost) 2 million regressions in the title of his paper result from counting the same models twice, since he distinguishes between identical models according to whether a variable is being "tested" or merely added in the triplet. See also his footnote 3.

probability of 0.81 percent, and it takes 20,274 models to account for 90 percent of the posterior mass. This makes a very strong case for conducting inference on the basis of BMA—as in (5)—rather than on one particular model. The 7,315 models in $\mathcal{M}_I$, corresponding to the most favored models in Sala-i-Martin (1997b) jointly obtain only 0.0021 percent of the total posterior mass. Interestingly, this is only one tenth of the prior mass assigned to $\mathcal{M}_I$, reflecting serious doubts about the adequacy of $\mathcal{M}_S$. In other words, the specific set of models with three retained regressors and exactly four additional regressors, considered in Sala-i-Martin, is not at all favored by the data. In fact, we find that all of the 45 models with posterior probabilities of at least 0.1 percent (which jointly account for 9.01 percent of the posterior mass) have in between 10 and 15 regressors.

Table 1 presents the posterior probabilities of including each of the 25 possible regressors.[5] A first observation is that the primary school enrollment rate is only included with probability 0.36, whereas it was always retained in $\mathcal{M}_S$ (such variables are preceded by a double arrow in the table). The probabilities of inclusion of the 22 preferred variables of Sala-i-Martin (1997b, Table 1) vary widely. Whereas some variables are almost always included (fraction Confucian, equipment in investment, sub-Saharan dummy, fraction Muslim, and fraction GDP in mining), quite a few have very low probabilities of inclusion (political rights, exchange rate distortions, Spanish colony, war dummy, latitude, and revolutions and coups). The rest are assigned probabilities in between 0.2 and 0.8. Thus, it is clear that the values of CDF(0), computed in Sala-i-Martin to be over 0.95 for each of these variables, are not closely related to the posterior probabilities of inclusion we find here.

Figure 1 graphically presents the posterior density functions of some key regression coefficients. A gauge on top of the graphs indicates (in black) the posterior probability of inclusion of the corresponding regressor. The actual posterior distribution is a mixture of the indicated continuous part and a point mass at zero, with the corresponding probabilities. The continuous part is itself a mixture as in (5) over the Student-t posteriors for each model. Note that this can lead to asymmetry (as in the coefficient of primary school enrollment) or even multimodality (see fraction Catholic). A dashed vertical line indicates the averaged point estimate presented in Tables 1 of Sala-i-Martin (1997a,b). Two vertical dotted lines indicate a classical 90 percent confidence interval[6] using the averaged estimated standard deviation of Sala-i-Martin (1997a,b). In spite of the difference in statistical paradigm,[7] Figure 1

[5] Defined as the sum of the posterior probabilities of all models that include this regressor.

[6] Sala-i-Martin's conclusions in the Normal case are based on whether or not this interval includes zero, and his results for the case with a non-Normal distribution of estimates across models are approximately based on this.

[7] In particular, we stress that our Bayesian analysis does not identify inclusion of a regressor with the scaled distance of the posterior mean from zero, but rather relies on explicit posterior

Table 1. Posterior Probabilities of Inclusion for Regressors
(Reduced set of regressors)

		Regressors	Post.Prob.
⇒	1	GDP level in 1960	1.0000
	2	Fraction Confucian	0.9997
⇒	3	Life expectancy	0.9975
	4	Equipment investment	0.9425
	5	Sub-Saharan dummy	0.8810
	6	Fraction Muslim	0.8734
	7	Fraction GDP in mining	0.8186
	8	Rule of law	0.7712
	9	Degree of capitalism	0.7368
	10	Nonequipment investment	0.6811
	11	Fraction Protestant	0.6796
	12	S.D. of black market premium	0.5602
	13	Fraction Buddhist	0.5502
	14	Number of years open economy	0.4599
⇒	15	Primary school enrollment, 1960	0.3562
	16	Primary exports, 1970	0.3273
	17	Civil liberties	0.2390
	18	Latin American dummy	0.2354
	19	Fraction Catholic	0.2140
	20	Political rights	0.1864
	21	Exchange rate distortions	0.1526
	22	Spanish Colony dummy	0.1508
	23	War dummy	0.1428
	24	Absolute latitude	0.0997
	25	Revolutions and coups	0.0974

Source: Authors' calculations.

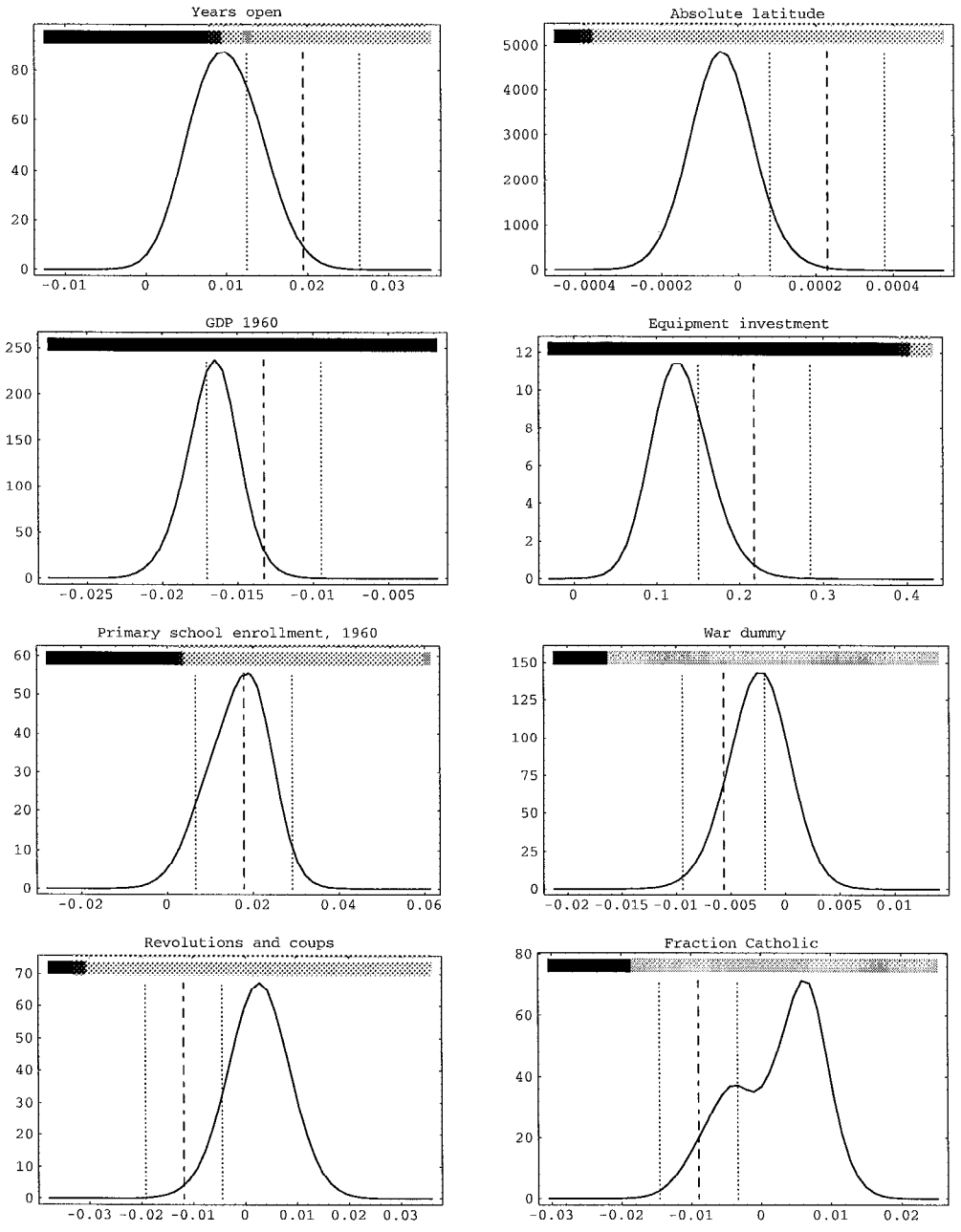
allows for an informal comparison with the findings of Sala-i-Martin concerning the inclusion of these regressors. For some coefficients, the main difference results from the location estimate (years open economy, latitude), for others (primary school enrollment, war) the spread is the main cause, and for fraction Catholic and revolutions and coups both location and spread are very different from the classical results in Sala-i-Martin.

In the above analysis, we have restricted ourselves to the 25 variables that were labeled as important in Sala-i-Martin (1997b). As our methodology—based on a much larger set of models—leads to quite different results, we might well question this choice of regressors. Thus, in order to tackle the problem more fully and propose a constructive alternative, based on our methodology, we now consider a larger set of possible regressors.

---

model probabilities. Nevertheless, the two often are correlated in practice.

Figure 1. Posterior Densities of Selected Coefficients
(Reduced set of regressors)



Source: Authors' calculations.

B. The Full Set of Regressors

We now add to the set in the previous subsection all regressors that do not entail a loss in the number of observations. Thus, we keep $n = 72$ observations which allows us to expand the set of regressors to a total of $k = 41$ possible variables. This implies we have a model space $\mathcal{M}$ with 2.2×10^{12} different models, and the task becomes daunting indeed. Again, we use a Uniform prior over models, and we find that (as expected, given the size of $\mathcal{M}$) we need a larger amount of drawings of the MC^3 sampler to achieve full convergence. We shall report results from a run with 2 million recorded drawings after a burn-in of 1 million, leading to a correlation coefficient of visit frequencies and actual posterior probabilities of 0.75. Nevertheless, even the results based on 500,000 drawings after a mere 100,000 burn-in drawings capture the essence of the posterior mass in $\mathcal{M}$. In particular, the best nine models are exactly the same in both runs. Many more runs, started from randomly drawn points in model space and leading to virtually identical results, confirmed convergence of the sampler. In a representative chain of 2 million draws, 522,476 models are visited (less than one in every four million models) and the best model obtains a posterior probability of 0.21 percent. Again, the mass is extremely spread out: the best 73,869 models share the first 90 percent of the posterior model probability, making BMA a necessity for meaningful inference. In addition, the nine best models all have in between 18 and 22 regressors, raising further doubt about the adequacy of the 7-regressor models in $\mathcal{M}_S$. Indeed, the intersection of $\mathcal{M}$ and $\mathcal{M}_S$, denoted by $\mathcal{M}_I$, is now allocated a mere 7.1×10^{-5} percent posterior probability (roughly 20 times the prior mass).[8] Note that $\mathcal{M}_I$ now contains 73,815 models.

Posterior probabilities of including each of the 41 regressors are tabulated in Table 2. In addition, the table contains the values of non-Normal, weighted CDF(0) that were at the basis of the findings of Sala-i-Martin (1997a,b). An arrow in front of a regressor identifies the 22 important regressors of Sala-i-Martin (1997b, Table 1) and regressors with double arrow are the ones he always retained in the models. Starting with the latter three, it is clear that life expectancy and GDP in 1960 can indeed be retained without any problem, but that is much less the case for primary school enrollment. For example, higher education enrollment would be a better choice from our results.

The 22 regressors that Sala-i-Martin flags as important have posterior probabilities of inclusion ranging from 4.8 percent (the lowest of all!) to 100.0 percent. The Spearman rank correlation coefficient between CDF(0) and posterior inclusion probabilities is only 0.29. Interestingly, many of the regressors get similar posterior

[8] For comparison, the posterior probability of the best 73,815 BMA models multiplies the corresponding prior probability by about 27 million.

Table 2. Posterior Probabilities of Inclusion for Regressors
(Full set of regressors)

		Regressors	BMA Post.Prob.	Sala-i-Martin CDF(0)
⇒	1	GDP level in 1960	1.000	1.000
→	2	Fraction Confucian	1.000	1.000
⇒	3	Life expectancy	0.999	0.999
→	4	Sub-Saharan dummy	0.996	0.997
→	6	Equipment investment	0.970	1.000
	5	Fraction Hindu	0.969	0.654
	7	Size labor force	0.945	0.835
→	8	Fraction GDP in mining	0.944	0.994
→	9	Rule of law	0.928	1.000
	10	Higher education enrollment	0.911	0.579
	11	Ethnolinguistic fractionalization	0.874	0.643
→	12	Latin American dummy	0.752	0.998
→	13	Nonequipment investment	0.714	0.982
→	14	Black market premium	0.685	0.825
⇒	15	Primary school enrollment, 1960	0.625	0.992
→	16	Spanish Colony dummy	0.593	0.938
→	17	Degree of capitalism	0.563	0.987
→	18	Civil liberties	0.545	0.997
	19	French Colony dummy	0.544	0.702
	20	British Colony dummy	0.468	0.579
→	21	Fraction Protestant	0.466	0.966
→	22	Fraction Muslim	0.439	1.000
	23	Outward orientation	0.418	0.634
	24	Fraction of pop. speaking English	0.416	0.910
	25	Public education share	0.297	0.580
→	26	Political rights	0.289	0.998
→	27	Fraction Buddhist	0.254	0.964
	28	Age	0.243	0.903
→	29	War dummy	0.143	0.984
→	30	Exchange rate distortions	0.120	0.968
→	31	Fraction Catholic	0.116	0.963
→	32	Number of years open economy	0.076	1.000
	33	Population growth	0.075	0.807
→	34	Absolute latitude	0.075	0.980
→	35	Primary exports, 1970	0.066	0.990
	36	Fraction speaking foreign language	0.058	0.831
	37	Area (scale effect)	0.057	0.532
	38	Ratio workers to population	0.053	0.766
→	39	S.D. of black market premium	0.050	0.993
	40	Fraction Jewish	0.048	0.747
→	41	Revolutions and coups	0.048	0.995

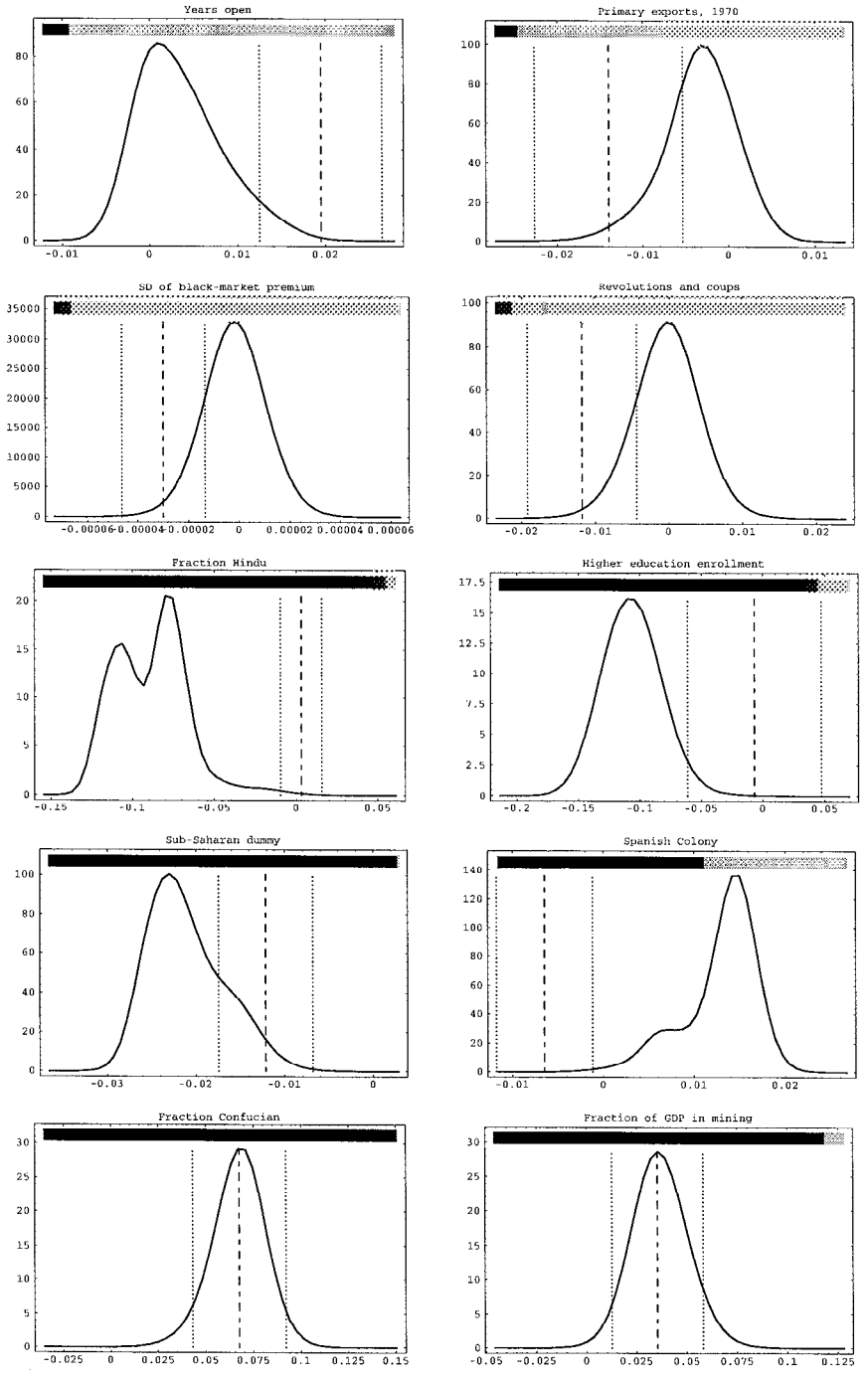
Source: Authors' calculations.

support for inclusion as in the analysis with the reduced set of regressors. In summary, our analysis leads to the classification shown in Table 2, where the top 11 regressors (with posterior probabilities larger than 0.85) could be labeled as *important* regressors, and the bottom 13 (with posterior probabilities smaller than 0.15) are *unimportant* regressors.

Thus, it is clear that our methodology does not lead to the same results as found in Sala-i-Martin (1997b). In addition, an important difference is that we have a coherent framework in which inference can be based on all models, averaged with their posterior probabilities. So there is no need for us to choose or discard any of the regressors, and the classification above need not be interpreted in that fashion. As we shall see in the next section, using BMA rather than choosing one particular model is particularly beneficial for prediction.

Figure 2 displays the posterior distributions of certain regression coefficients in the same format as Figure 1. The first four coefficients graphed in Figure 2 correspond to four variables—years open economy, primary exports, standard deviation of black market premium, and revolutions and coups—that are each included in Sala-i-Martin's analysis (with CDF(0) $\geq$ 0.990), yet get less than 8 percent posterior probability of inclusion in our BMA analysis. Averaged posterior modes are quite far from the averaged classical point estimates. The next two coefficients are illustrative of the opposite situation. The regressors fraction Hindu and higher education enrollment are excluded in Sala-i-Martin (1997b) with CDF(0) $<$ 0.66, but BMA results indicate these are quite important variables with posterior probabilities of inclusion exceeding 91 percent. Two regressors for which results of both analysis are seemingly similar are the dummies for sub-Saharan country and Spanish colony. For the former, BMA inclusion probability and CDF(0) are both very close to unity, yet the posterior mode is roughly twice the averaged point estimate. For the latter, both analyzes might well agree on the relative importance of the variable, but the sign and magnitude of the effect of this variable differ dramatically between the two methods. Finally, for fraction Confucian and fraction of GDP in mining the averaged confidence intervals accord well with the posterior results from BMA.

Figure 2. Posterior Densities of Selected Coefficients
(Full set of regressors)



Source: Authors' calculations.

IV. PREDICTIVE RESULTS

An important quality of a model is that it can provide useful forecasts. In addition, such a predictive exercise immediately provides a benchmark for evaluating the model's adequacy. We consider predicting the observable y_f given the corresponding values of the regressors, grouped in a k-dimensional vector z_f (which has been transformed in the same way as Z, by subtracting the average of the original regressors over the n observations on which posterior inference is based.[9]

Prediction naturally fits in the Bayesian paradigm as all parameters can be integrated out, formally taking parameter uncertainty into account. If we also wish to deal with model uncertainty, BMA as in (5) provides us with the formal mechanism, and we can characterize the out-of-sample predictive distribution of y_f by

$$\begin{aligned} p(y_f \mid y) = \sum_{j=1}^{2^k} & f_S(y_f \mid n-1, \overline{y} + \frac{1}{g_{0j}+1} z'_{f,j}\beta^*_j, \\ & \frac{n-1}{d^*_j}\{1 + \frac{1}{n} + \frac{1}{g_{0j}+1} z'_{f,j}(Z'_jZ_j)^{-1}z_{f,j}\}^{-1})P(M_j \mid y), \end{aligned} \tag{9}$$

where $f_S(x \mid \nu, b, a)$ denotes the p.d.f. of a univariate Student-t distribution with ν degrees of freedom, location b (the mean if $\nu > 1$) and precision a (with variance $\nu/\{a(\nu-2)\}$ provided $\nu > 2$) evaluated at x. In addition, $z_{f,j}$ groups the j elements of z_f corresponding to the regressors in M_j, $\beta^*_j = (Z'_jZ_j)^{-1}Z'_jy$ and

$$d^*_j = \frac{1}{g_{0j}+1} y'M_{X_j}y + \frac{g_{0j}}{g_{0j}+1}(y - \overline{y}\iota_n)'(y - \overline{y}\iota_n). \tag{10}$$

The term in (9) corresponding to the model with only an intercept is obtained by letting the corresponding g_{0j} tend to infinity.

We shall now split the sample into n observations on which we base our posterior inference and q observations which we retain in order to check the predictive accuracy of the model. As a formal criterion, we shall use the log predictive score, introduced by Good (1952). For $f = n+1, \ldots, n+q$ (*i.e.*, for each country in the prediction sample) we base our predictive measure on (9) evaluated in these retained observations $y_{n+1}, \ldots, y_{n+q}$, namely:

$$LPS = -\frac{1}{q}\sum_{f=n+1}^{n+q} \ln p(y_f \mid y), \tag{11}$$

[9] This is merely to assign the same interpretation to the regression coefficients in posterior and predictive analysis.

The smaller LPS is, the better the model does in forecasting the prediction sample.

In particular, we shall compare four different regression models: the BMA model, given by (9), the best model (*i.e.*, the one with the highest posterior probability) in $\mathcal{M}$, the best model in $\mathcal{M}_I$, and the full model with all k regressors. As a benchmark for the importance of growth regression, we also include LPS for the "null model," *i.e.*, the model with only the intercept where no individual country characteristics are used in predicting the growth of the countries in the prediction sample.

The partition of the sample into the inference and the prediction sample is done randomly, by assigning an observation to the inference sample with probability 0.75. The results of three different partitions are presented in Table 3, for the reduced set of regressors ($k = 25$) and the full set of regressors ($k = 41$). Results cannot be compared across partitions, but the comparison between the different regression models clearly favors BMA. In particular, BMA dominates the single regression models in five out of the six cases examined, often by a large margin. In one case, the best model from the set $\mathcal{M}_I$ does slightly better than BMA. Overall, this is very compelling evidence supporting the use of formal model averaging rather than the selection of any given model. Finally, from the predictive performance of the null model, we deduce that growth regression through BMA typically improves predictions dramatically, except for one case. In the first run of the case with $k = 41$ the null model proved to be best, undoubtedly as a result of the fact that the countries in the estimation sample were not at all representative for those in the prediction sample, and our regression inference thus actually led us astray (but, still, least of all for BMA). Overall, however, regression models with BMA result in a considerable predictive improvement over the null model.

Table 3. Predictive Performance

	$k = 25$			$k = 41$		
	Run 1	Run 2	Run 3	Run 1	Run 2	Run 3
BMA	0.80	0.35	1.67	4.49	0.10	1.94
Best model in $\mathcal{M}_I$	1.21	2.06	3.42	5.12	0.32	1.54
Full model	3.11	0.96	2.00	7.57	0.22	4.24
Best model in $\mathcal{M}$	4.52	3.36	2.84	9.65	2.83	3.94
Null model	2.69	2.83	2.51	2.24	2.67	2.48
q	20	21	20	18	23	18

Source: Authors' calculations.

V. DISCUSSION

The value of growth regression in cross-country analysis has been illustrated in the predictive exercise in the previous section. We agree with Sala-i-Martin (1997b) that some regressors can be identified as useful explanatory variables for growth in a linear regression model, but we advocate a formal treatment of model (and parameter) uncertainty. In our methodology the importance of an explanatory variable does not necessarily imply anything about the size or sign of the regression coefficient in a set of models, but is based entirely on the posterior probabilities of models containing that regressor. In addition, we go one step further and provide a practical and theoretically sound method for inference, both posterior and predictive, using BMA. 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 regression.

Our Bayesian paradigm provides us with a formal framework to implement this model averaging, and recent MCMC methods are shown to be very powerful indeed. Despite the huge model space (with 2.2 trillion models), we obtain reliable results without an inordinate amount of computational effort.[10]

The analysis in Sala-i-Martin (1997b) is not Bayesian and thus no formal model averaging can occur, even though he considers weighting with the integrated likelihood.[11] In addition, the latter analysis evaluates all models and is thus necessarily restricted to a rather small set of models, $\mathcal{M}_S$, which turns out not to receive much empirical support from the data. As a consequence, we find a rather different set of variables that can be classified as "important" for growth regressions. An important additional advantage is that our results are immediately interpretable in terms of model probabilities and all inference can easily be conducted in a purely formal fashion by BMA. It is not clear to us what to make of the recommendations in Sala-i-Martin (1997b): should the applied researcher use all of the regressors identified as important or mix over the corresponding models in $\mathcal{M}_S$? However,

[10] The chain for the reduced set of regressors takes about 1 hour and 45 minutes on a 240MHz PowerPC 603-based laptop (Apple PowerBook 3400c under MacOS 8.5, with 144Mb of physical RAM, 256Mb of total RAM). Each posterior density for the coefficients takes about one additional minute to compute. The computation of LPS takes about three seconds for each prediction involved. Computation times are cut, at least, in half in a shared Sun SPARCenter 1000E under SunOS 5.4 with 3.5Gb of RAM. The chains for the full set of 41 regressors were all executed in the Sun SPARCenter. The reported chain took about 21 hours of CPU, and each coefficient posterior density about five additional minutes. For the 75-25 prediction split, the computation of LPS required about 30 seconds for each prediction. Our programs are efficiently coded in Fortran 77 and are freely available at http://www.freeyellow.com/members6/mcmcmc/.

[11] Sala-i-Martin (1997a,b) does not specify what this integrated likelihood is; as there is no prior to integrate with, this may refer to the maximized likelihood, which is proportional to $(y'M_{X_j}y)^{-n/2}$ for M_j.

the latter would have to be without proper theoretical foundation or guidance if a classical statistical framework is adopted.

In our view, the treatment of a very large model set, such as $\mathcal{M}$, in a theoretically sound and empirically practical fashion requires BMA and MCMC methods. In addition, this methodology provides a clear and precise interpretation of the results, and immediately leads to posterior and predictive inference.

REFERENCES

Barro, Robert J., 1991, "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, Vol. 106, pp. 407–44.

Chib, Siddhartha, and Edward Greenberg, 1996, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, Vol. 12, pp. 409–31.

Durlauf, Steven N., and Danny T. Quah, 1999, "The New Empirics of Economic Growth," in *Handbook of Macroeconomics*, ed. by J. Taylor and M. Woodford (New York: North-Holland).

Fernández, Carmen, Eduardo Ley, and Mark F.J. Steel, 1998, "Benchmark Priors for Bayesian Model Averaging," FEDEA Working Paper No. 98-06 (Madrid: Fundación de Estudios de Economía Aplicada).

Good, I.J., 1952, "Rational Decisions," *Journal of the Royal Statistical Society*, Series B, Vol. 14, pp. 107–14.

Kormendi, Roger, and Philip Meguire, 1985, "Macroeconomic Determinants of Growth: Cross-Country Evidence," *Journal of Monetary Economics*, Vol. 16, pp. 141–63.

Leamer, Edward E., 1978, *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (New York: Wiley).

———, 1983, "Let's take the Con out of Econometrics," *American Economic Review*, Vol. 73, pp. 31–43.

———, 1985, "Sensitivity Analyses Would Help," *American Economic Review*, Vol. 75, pp. 308–13.

Lee, Herbert, 1996, "Model Selection for Consumer Loan Application Data," Department of Statistics Working Paper No. 650 (Pittsburgh, Pennsylvania: Carnegie-Mellon University).

Levine, Ross, and David Renelt, 1992, "A sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, Vol. 82, pp. 942–63.

Madigan, David, and J. York, 1995,"Bayesian Graphical Models for Discrete Data," *International Statistical Review*, Vol. 63, pp. 215–32.

Raftery, A.E., David Madigan, and Jennifer A. Hoeting, 1997, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, Vol. 92, pp. 179–91.

Sala-i-Martin, Xavier, 1997a, "I Just Ran Four Million Regressions" (unpublished; New York, New York: Columbia University).

———, 1997b,"I Just Ran Two Million Regressions," *American Economic Review*, Vol. 87, pp. 178–83.

Temple, Jonathan, 1999, "The New Growth Evidence," *Journal of Economic Literature*, Vol. 37, pp. 112–56.