DETERMINANTS OF LONG-TERM GROWTH: A BAYESIAN AVERAGING OF CLASSICAL ESTIMATES (BACE) APPROACH

Gernot Doppelhofer
Ronald I. Miller
Xavier Sala-i-Martin

Determinants of Long-Term Growth: A Bayesian Averaging of
Classical Estimates (BACE) Approach
Gernot Doppelhofer, Ronald I. Miller, and Xavier Sala-i-Martin
NBER Working Paper No. 7750
June 2000
JEL No. O51, O52, O53

ABSTRACT

This paper examines the robustness of explanatory variables in cross-country economic growth regressions. It employs a novel approach, Bayesian Averaging of Classical Estimates (BACE), which constructs estimates as a weighted average of OLS estimates for every possible combination of included variables. The weights applied to individual regressions are justified on Bayesian grounds in a way similar to the well-known Schwarz criterion. Of 32 explanatory variables we find 11 to be robustly partially correlated with long-term growth and another five variables to be marginally related. Of all the variables considered, the strongest evidence is for the initial level of real GDP per capita.

Gernot Doppelhofer
Department of Economics
Columbia University
420 West 118th Street, 1005
New York, NY 10027
and Cambridge University

Xavier Sala-I-Martin
Department of Economics
Columbia University
420 West 118th Street, 1005
New York, NY 10027
and NBER, and Universitat Pompeu Fabra
xs23@columbia.edu

Ronald I. Miller
Department of Economics
Columbia University
420 West 118th Street, 1005
New York, NY 10027

TABLE OF CONTENTS

1.- INTRODUCTION

Following the seminal work of Barro (1991), the recent empirical literature on economic growth has identified a substantial number of variables that are partially correlated with the rate of economic growth. The basic methodology consists of running cross-country regressions of the form[1]

$$\gamma = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon \tag{1}$$

where γ is the vector of rates of economic growth, and $x_1,...,x_n$ are vectors of explanatory variables which vary across researchers and across papers. Each paper typically reports a (possibly non-random) sample of the regressions actually run by the researcher. Variables like the initial level of income, the investment rate, various measures of education, some policy indicators and many other variables have been found to be significantly correlated with growth in regressions like (1).

The problem faced by empirical growth economists is that growth theories are not explicit enough about what variables x_j belong in the "true" regression. That is, even if we know that the "true" model looks like (1), we do not know exactly what variables x_j we should use. One reason is that economic growth theory is not explicit about what variables matter for growth. For example, almost all growth theories say that the "level of technology" [the constant "A" in the typical production function, $Y=F(K,L,A)$] is an important determinant of growth, at least along a

---

[1] Recently, a number of authors have broken up the period of analysis into various sub-periods and have estimated the same type of regressions using panel techniques. See for example, Islam (1995), Caselli, Esquivel, and Laffort (1996) or Barro and Sala-i-Martin (1995).

transition towards the steady state.[2]   From a macroeconomic perspective, there are many things

other than the "engineering" level of technology which can be thought of as "the level of

technology," A. In other words, a lot of factors may affect the aggregate amount of output, given

the aggregate amount of inputs. These may include market distortions, distortionary taxes,

maintenance of property rights, degree of monopoly, weather, attitudes toward work, and so on.

Hence, creative theorizing will generate models that "predict" that any of these or other variables

should be included in the growth regression.

The multiplicity of possible regressors is one of the major difficulties faced by

researchers trying to make sense of the empirical evidence on economic growth. However, the

problem is hardly unique to the growth literature: artistic economic theory is often capable of

suggesting an enormous number of potential explanatory variables in any economic field. In

principle, this is strictly a small-sample problem since, as the number of observations becomes

large, all of the variables which do not belong in the regression will have coefficients that

converge to zero. Thus, classical statistics offers us little help: we should simply include all of

the suggested regressors and let the data sort through them. When questions can be addressed

with very large datasets it is routine practice to include every regressor that comes to mind and

then report those that have significant coefficients. Often, however, we do not have the luxury

of having a sample size that allows us to include all potential regressors. Cross-country growth

regressions provide perhaps the most extreme example: the number of proposed regressors

---

[2] Theories of endogenous growth suggest that such constant is a determinant of the steady-state
growth rate while neoclassical models argue that this is true along the transition only. Our
argument is completely independent of such disputes.

exceeds the number of countries in the world, rendering the all-inclusive regression computationally impossible.

The methodology usually used by empirical economists consists on simply "trying" the variables which are thought to be potentially important determinants of growth. However, as soon as one starts running regressions combining the various variables, one finds that variable x_1 is significant when the regression includes variables x_2 and x_3, but it becomes insignificant when x_4 is included. Since one does not know *a priori* the "true" variables to be included, one is left with the question: what variables are "truly" correlated with growth?

An initial answer to this question was given by Levine and Renelt (1992). They applied a modified[3] version of Leamer's (1983, 1985) *extreme bounds analysis* to identify "robust" empirical relations for economic growth. In short, the extreme bounds test works as follows: imagine that we have a pool of K variables previously identified as related to growth and are interested in knowing whether variable z is "robust." We would estimate regressions of the form:

$$\gamma = \alpha_j + \beta_{yj} \cdot y + \beta_{zj} \cdot z + \beta_{xj} \cdot x_j + \varepsilon \tag{2}$$

where y is a vector of *fixed* variables that always appear in the regressions (in the Levine and Renelt paper, these variables are the *initial level of income*, the *investment rate*, the *secondary school enrollment rate* and the *rate of population growth*), z is the variable of interest and x_j is

---

[3] We say "modified" because they limited the number of regressors to be included in each regression as opposed to the original Leamer technique which allows all potential combinations of regressors.

a vector of up to three variables taken from the pool of the K variables available. One needs to estimate this regression or model for all possible x_j combinations. For each model, j, one finds an estimate, $\hat{\beta}_{zj}$, and the corresponding standard deviation, $\hat{\sigma}_{zj}$. The *lower extreme bound* is defined to be the lowest value of $\hat{\beta}_{zj} - 2\hat{\sigma}_{zj}$ over all possible models j, and the *upper extreme bound* is defined to be the largest value of $\hat{\beta}_{zj} + 2\hat{\sigma}_{zj}$. The *extreme bounds test* for variable z says that if the lower extreme bound is negative and the upper extreme bound is positive, then variable z is fragile.

Not surprisingly, Levine and Renelt's conclusion is that very few (or no) variables are robust. One possible reason for finding few or no robust variables is, of course, that very few variables can be identified as correlated systematically with growth. Hence, some researchers have been tempted to conclude that *"nothing can be learned from this empirical growth literature because no variables are robustly correlated with growth."* Another interpretation, however, is that the test is too strong for any variable to pass: if there is *one* regression for which the sign of the coefficient β_z changes, or becomes insignificant, then the variable is labeled as "fragile." This is independent of how poorly the regression fits: all regressions are treated equally and the statement of any one of them carries a veto.[4] This problem is well recognized and some solutions have been proposed such as the *reasonable extreme bounds* of Granger and Uhlig (1990).[5]

---

[4] There are other criticisms of extreme bounds analysis; see for example Durlauf and Quah (1999).

[5] See Doppelhofer (2000) for an application of Granger-Uhlig's *reasonable extreme bounds* to cross-country growth regressions.

Sala-i-Martin (1997a and b) proposes to depart from this "extreme" test and, instead of assigning a label of "fragile" or not to a particular variable, he decides to assign some "level of confidence" to each variable. To this end, he constructs weighted averages of all the estimates of $\hat{\beta}_{zj}$ and its corresponding standard deviations, $\hat{\sigma}_{zj}$, using weights proportional to the likelihoods of each of the models. As a measure of significance Sala-i-Martin calculates a likelihood-weighted sum of normal cumulative distribution functions. He finds that Levine and Renelt's pessimistic conclusion is not warranted and that a number of variables are significantly correlated with growth. In order to maintain comparability, Sala-i-Martin follows Levine and Renelt in assuming that there is a set of "fixed regressors" which belong in all models[6], and he restricts all the regressions to have the same size of seven regressors.

A natural way to think about model uncertainty, related to Sala-i-Martin's approach, is to admit that we do not know which model is "true" and, instead, attach probabilities to different possible models. While intuitively appealing, this requires a departure from the classical framework in which conditioning on a model is essential. This approach has recently come to be known as *"Bayesian Model Averaging"*. The procedure does not differ from the most basic Bayesian reasoning: the idea dates at least to Jeffreys (1961) although fleshed out by Leamer

---

[6] The fixed regressors in Sala-i-Martin are the initial level of income per capita, the life expectancy and primary school enrollment in 1960. Even though he checks the significance of these three variables, the computed "model averages" all include these three variables, which may be problematic, especially if some of the variables tested are highly correlated with the variables that are always included.

(1978). In this paper, we show that this approach can be used in a way that is well grounded in statistical theory, intuitively appealing, easy to understand, and easy to implement.[7]

The fully Bayesian approach is entirely feasible and has been applied to various problems by a number of authors. Examples include Raftery, Madigan and Hoeting (1997) and York, Madigan, Heuch and Lie (1995).[8] In the growth context, Fernandez, Ley and Steel (2000) apply techniques from the Bayesian statistics literature to the dataset of Sala-i-Martin (1997a). A pure Bayesian approach requires specification of the prior distributions of all of the relevant parameters *conditional on each possible model*.[9] Under ideal conditions, elicitation of prior parameters is difficult and is indeed one of the major reasons for Bayesian approaches remaining relatively unpopular. But when the number of possible regressors is K, the number of possible linear models is 2^K so with K large, fully specifying priors is infeasible. Thus, authors implementing the fully Bayesian approach have used priors which are essentially arbitrary. This makes the ultimate estimates dependent on arbitrarily chosen prior parameters in a manner which is extremely difficult to interpret. In existing applications of this approach, the impact of these prior parameters has been neither examined nor explained.

In this paper we will use the Bayesian approach to averaging across models, while following the classical spirit of Sala-i-Martin (1997a and b). We propose a model-averaging

---

[7] Although the computational burden of our procedure is not insignificant it can be executed on a current PC.

[8] A summary of much of the recent work can be found in Hoeting, Madigan, Raftery and Volinsky (1999).

[9] For readers unfamiliar with Bayesian language, the *prior distribution*, is a summary of the researchers beliefs concerning the parameters, prior to seeing the data.

technique which we call *Bayesian Averaging of Classical Estimates* or *BACE*, to determine the

"importance" of variables in cross-country growth regressions. We show that the weighting

method can be derived as a limiting case of a standard Bayesian analysis as the prior information

becomes "dominated" by the data. BACE combines the averaging of estimates across models,

which is a Bayesian concept, with Classical OLS estimation which comes from the assumption of

diffuse priors. This name is chosen to highlight the fact that while averaging across models is an

inherently Bayesian idea, BACE limits the effect of prior information and uses an approach

otherwise familiar to classical econometricians.

Our BACE approach has several important advantages over previously used model-

averaging and robustness-checking methods: firstly, in contrast to a standard Bayesian approach

that requires the specification of a prior distribution for all parameters, BACE requires the

specification of only one prior hyper-parameter, the expected model $\bar{k}$ size.   This parameter is

easy to interpret, easy to specify, and easy to check for robustness.[10] Secondly, the interpretation

of the estimates is straightforward for economists not trained in Bayesian inference. The weights

applied to different models are proportional to the logarithm of the likelihood function corrected

for degrees of freedom (analogous to the Schwarz model selection criterion). Thirdly, our

estimates can be calculated using only repeated applications of OLS. Fourthly, in contrast to

Levine and Renelt and Sala-i-Martin, we consider models of all sizes and no variables are held

"fixed" and therefore "untested."

_____

[10]In the standard Bayesian sense that we can calculate estimates for a range of different values of
$\bar{k}$. Thus we can make statements of the form, "whether you think a good model size is three
regressors or 12 regressors, this one particular variable is important".

When we examine the cross-country data usually used by growth empiricists using this approach we find striking and surprisingly clear conclusions. The data identify a set of four variables which have a great deal of explanatory power and are very precisely estimated. A second group, with seven variables, has somewhat weaker explanatory power, but have coefficient which are still quite precisely estimated. Another five variables are marginal: they would be reasonable regressors if a researcher had a strong prior belief in their relevance. The remaining sixteen variables have weak explanatory power and are imprecisely estimated.

The rest of the paper is organized as follows. In Section 2 we outline the statistical theory in which our estimates tests are based. In Section 3 we describe the data set used. Section 4 presents the main empirical results of the paper. The final section concludes.

2.- STATISTICAL THEORY

2.1. Statistical Basics

Following is a quick exposition of the basic reasoning and the language needed for understanding our approach. An extremely clear and more detailed presentation of these ideas can be found in Poirier (1995). We begin with Bayes' rule. This is simply a theorem, a basic consequence of conditional probability. Bayes' rule in densities is:

$$g(\beta|y) = \frac{f(y|\beta)g(\beta)}{f(y)} \qquad \textbf{(3)}$$

This is true for any random variables y and β. In classical statistics a parameter has a true, though unknown, value, so it cannot have a density because it is not random. In the Bayesian framework

parameters are considered to be uncertain. In (3) above, $g(\beta)$ is the *prior density* of a parameter

vector β, interpreted as the researcher's information about β *prior* to seeing the data. The vector

y is the observed data and *f(y)* is its density. The left-hand side of (3), $g(\beta|y)$, is the density of β

conditional on the data and is called the *posterior density*: it fully describes what a Bayesian

researcher knows about the parameters after seeing the data. Thus, in a Bayesian interpretation,

Bayes' rule tells us how to combine prior information with new data in order to get our final

opinions, or posterior beliefs.

"Model averaging" is a special case of Bayes' rule. Suppose we divide the parameter

space into two regions and label them M_0 and M_1. These regions could be what we would

usually call hypotheses (e.g., $\beta > 0$ versus $\beta \leq 0$) or something we would usually call models (e.g.,

$\beta_1 = 0, \beta_2 \neq 0$ versus $\beta_1 \neq 0, \beta_2 = 0$). Each of these has a prior probability specified by the

researcher as $P(M_i)$. These prior probabilities summarize the researcher's beliefs concerning the

relative likelihood of the two regions (models). Given the two regions, Bayes' rule implies

$g(\beta|y) = P(M_0)\dfrac{f(y|\beta)g(\beta|M_0)}{f(y)} + P(M_1)\dfrac{f(y|\beta)g(\beta|M_1)}{f(y)}$. Rewriting this in terms of the posterior

probabilities conditional on the two regions (models) we get:

$$g(\beta|y) = P(M_0|y)\frac{f(y|\beta)g(\beta|M_0)}{f(y|M_0)} + P(M_1|y)\frac{f(y|\beta)g(\beta|M_1)}{f(y|M_1)} \qquad (4)$$

where $\quad P(M_i|y) \quad$ is the posterior probability of the i'th region, the probability of that region

conditional on the data. In words, equation (4) says that the posterior distribution of the

parameters is the weighted average of the two possible conditional posterior densities with the

weights given by the posterior probabilities of the two regions. In this paper we will be

considering linear regression models for which each model is a list of included variables, with the slope coefficients for all of the other possible regressors set equal to zero.

Much of the Bayesian statistics literature consists of formulae and methods for calculating the various quantities in equation (4) for different statistical models. For the linear regressions models examined here we will be able to refer to textbook derivations. The difficult part lies in deriving the posterior model probabilities.

2.2. Diffuse Priors

As noted, fully specifying priors is infeasible when the set of possible regressors is large. In applications of Bayesian theory, if a researcher is incapable or unwilling to specify prior beliefs a standard remedy is to apply *diffuse priors*. Though there are some difficulties with this notion, it is one way to represent initial ignorance. If the parameter space is bounded then a diffuse prior is a uniform distribution. When the parameter space is unbounded, as in the usual multiple linear regression model, a uniform distribution cannot be directly imposed and instead we must take a limit as the prior distribution becomes flat. In many contexts, imposing diffuse priors generates classical results: in the linear regression model standard diffuse priors and Bayesian regression yields posterior distributions identical to the classical sampling distribution of OLS.

We would like to work with diffuse priors but they create a problem when different regression models contain different sets of variables. As noted above, when the parameter space is unbounded we must get results for diffuse priors by taking a limit of informative priors. The informative prior must specify prior information concerning both β, the vector of slope

coefficients, and σ, the error standard deviation. There are no difficulties taking the limit as our prior information concerning σ becomes uninformative so the equations below all reflect a diffuse prior with respect to σ.   Equation (5) below gives the ratio of the posterior probabilities of two regression models (called the posterior odds ratio) with different sets of included variables, X for M_0 and Z for M_1.

$$\frac{P(M_0|y)}{P(M_1|y)} = \frac{P(M_0)}{P(M_1)} \left(\frac{|A|/|A+X'X|}{|B|/|B+Z'Z|} \right)^{-1/2} \left(\frac{SSE_0 + Q_0}{SSE_1 + Q_1} \right)^{-T/2} \tag{5}$$

where $P(M_i)$ is the prior probability of model i as specified by the researcher. This expression assumes that the marginal prior density for β is multivariate normal with variance-covariance matrices given by A^{-1} under M_0, and by B^{-1} under M_1. SSE_i is the OLS sum of squared errors under model i, T is the sample size and Q_i is a quadratic form in the OLS estimated parameters that need not concern us here. This is a textbook expression (e.g., Zellner (1971)). Making the priors diffuse requires taking the limit of (5) as A and B approach zero so that the variance of our prior density goes to infinity. The mathematical difficulty with this is the factor in (5) with the ratio of the determinants of A and B.   Both determinants approach zero as the variance goes to infinity, so their ratio depends on the *rate* at which each goes to zero. Depending on precisely how one parameterizes the matrices one gets different answers when evaluating this limit.[11]   One limit is the likelihood-weighting method of Sala-i-Martin (1997a). If we specify the prior

---

[11] Leamer (1978) provides some intuition for why such problems occur but argues, in Bayesian spirit, that one should not be interested in diffuse priors.

precision matrices as $A = gX'X$ and $B = gZ'Z$ (Zellner's (1986) g-prior) and take the limit of (5) as g goes to zero we get:

$$\frac{P(M_0|y)}{P(M_1|y)} = \frac{P(M_0)}{P(M_1)} \left(\frac{SSE_0}{SSE_1} \right)^{-T/2} \tag{6}$$

The second factor on the right-hand side is equal to the likelihood ratio of the two models. This weighting is troublesome because models with more variables have lower SSE's; the posterior mean model size (average of the different model sizes weighted by their posterior probabilities) will be bigger than the prior, whatever the data that is actually seen. Thus it is not sensible to use this approach when considering models of different sizes.

The indeterminacy of the limit in (5) suggests that for fairly diffuse priors the exact specification of the prior precision matrix, which will in practice be arbitrary, may generate large differences in the results. There is, however, another limit one can take: the limit of (5) as the information in the data, $X'X$, becomes large. The idea here is we are taking the limit as the prior becomes "dominated" by the data. Instead of taking the limit as the prior becomes flat we are taking the limit as the data becomes very informative relative to the prior information. If we assume that the variance-covariance matrix of the X's exists and take the limit of (5) as $X'X$ goes to infinity we get:[12]

---

[12] See Leamer (1978) for a similar expression.

$$\frac{P(M_0|y)}{P(M_1|y)} = \frac{P(M_0)}{P(M_1)} \, T^{(k_1 - k_0)/2} \left(\frac{SSE_0}{SSE_1} \right)^{-T/2} \tag{7}$$

where k_i is the number of included regressors in model M_i.[13] This provides an approximation to

the odds ratios generated by a wide range of reasonably diffuse prior distributions. The degrees-

of-freedom correction should be familiar, since it is the ratio of the Schwarz criteria for the two

models, exponentiated. The similarity to the Schwarz criterion is not coincidental: Schwarz

(1978) used the same approximation to the odds ratio to justify the criterion. In our empirical

work we will use the approximation in equation (7).

In order to get weights for different models we need the posterior probabilities of each

model, not the odds ratio. However, using the odds ratio given by (7), to get an approximate

posterior model probability we simply need to normalize the weight of a given model by the sum

of the weights of all possible models, i.e., with K possible regressors:

$$P(M_j|y) = \frac{P(M_j) \, T^{-k_j/2} SSE_j^{-T/2}}{\sum_{i=1}^{2^K} P(M_i) \, T^{-k_i/2} SSE_i^{-T/2}} \tag{8}$$

Once the model weights have been calculated, Bayes' rule says that the posterior density of a

parameter is the average of the posterior densities conditional on the models as shown in (4) for

---

[13]This precise expression arises only if we take the limit using g-priors. For other sorts of priors
it is an approximation.

two models. A posterior mean is defined to be the expectation of a posterior distribution.

Taking expectations with respect to β across (4) (with 2^K terms instead of only two) gives:

$$E(\beta|y) = \sum_{j=1}^{2^K} P(M_j|y)\hat{\beta} \tag{9}$$

where $\hat{\beta}_j = E(\beta|y, M_j)$ is the OLS estimate for β with the regressor set that defines model j. In

Bayesian terms, $\hat{\beta}_j$ is the posterior mean conditional on model J.[14] Note that any variable

excluded from a particular model has a slope coefficient with degenerate posterior distribution at

zero. The posterior variance of β is given by:

$$Var(\beta|y) = \sum_{j=1}^{2^K} P(M_j|y)Var(\beta|y, M_j) + \sum_{j=1}^{2^K} P(M_j|y)\left(\hat{\beta}_j - \sum_{j=1}^{2^K} P(M_j|y)\hat{\beta}_j\right)^2 \tag{10}$$

Leamer (1978) provides a simple derivation for (10). Inspection of (10) demonstrates that the

posterior variance incorporates both the estimated variances in individual models as well as the

variance in estimates of the β's across different models.

While posterior means and variances are certainly of interest, there are other ways to

summarize the large amount of information supplied by the full posterior distribution. In

particular we would like to know the posterior probability that a particular variable is in the

regression (i.e., has a non-zero coefficient). We will call this the *posterior inclusion probability*

---

[14] The difficulty with making the prior diffuse applies only to the comparison, or averaging, of
different models. Conditional on one particular set of included variables the mean of the
Bayesian regression posterior is simply the OLS estimate.

for the variable and it is calculated as the sum of the posterior model probabilities for all of the models including that variable. We will also report the posterior mean and variance conditional on the inclusion of the variable.

2.3. Issues in Implementation

2.3.1. Model Size

We have not yet discussed the specification of the $P(M_i)$'s, the prior probabilities attached to the different models. One common approach to this problem in the statistical literature has been to assign equal prior probability to each possible model. While this is sensible for some applications, for linear regression with a large number of potential regressors it has odd and troubling implications. In particular it implies a very strong prior belief that the number of included variables should be large. We will instead specify our model prior probabilities by choosing a prior mean model size, $\bar{k}$, with each variable having a prior probability $\bar{k}/K$ of being included, independent of the inclusion of any other variables, where K is total number of potential regressors.[15] Equal probability for each possible model is the special case in which $\bar{k}=K/2$. In our empirical work we focus on a relatively small $\bar{k}$ on the grounds that most researchers prefer relatively modest parameterizations. We examine the robustness of our conclusions with respect to this hyper-parameter in Section 4.2.

---

[15] In most applications the prior probability of including a particular variable is not, for most researchers, independent of the probability of including any other variable. For example, in a growth regression if a variable proxying political instability is included, such as a count of revolutions, many researchers would be think it less likely that another measure such as the number of assassinations be included. While this sort of interdependence can be readily incorporated into our framework, we do not presently pursue this avenue.

Prior Probabilities by Model Size: Kbar =7

Prior Probabilities By Model Size: Equal Model Probabilities (Kbar = 16)
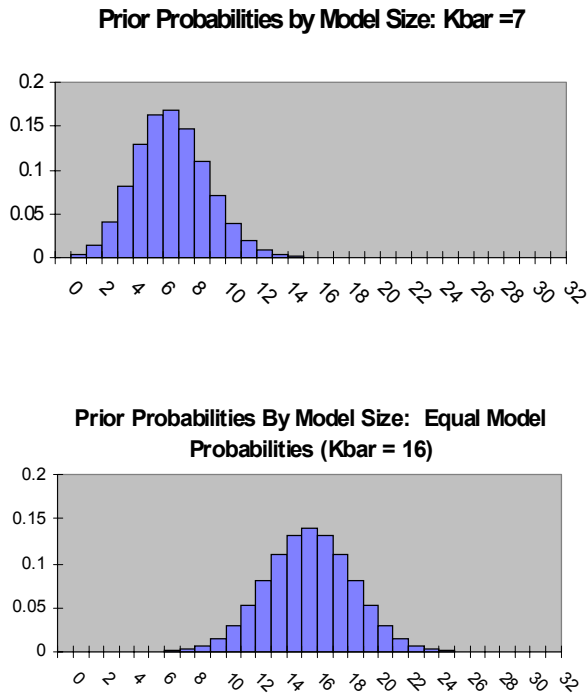
In order to illustrate further this issue the two figures shown graph the prior probability distribution by model size for our baseline model with $\bar{k}=7$ and with equal probabilities for all models, $\bar{k}=16$, given the 32 potential regressors we consider in our empirical work. Note that in the second chart, the great majority of the prior probability is focused on models with many included variables: more than 99% of the prior probability is located in models with ten or more included variables. It is our strong opinion that few researchers actually have such prior beliefs. Thus while we will calculate results for the equal model probability case below, we do not choose to focus attention on this case.

2.3.2. Sampling

Equations (8), (9) and (10) all face the problem that they include sums running over 2^K terms: for many problems for which model averaging is attractive this is an infeasibly large number even though each term only requires the computation of an OLS regression. For our baseline estimation, with $k = 32$, this is around 4.3 billion regressions. As a result, only a relatively small subset of the total number of possible regressions can be run.

17

Several stochastic algorithms have been proposed for dealing with this issue, including the Markov-Chain Monte-Carlo Model Composition (MC^3) technique (Madigan and York (1995)), SSVS (George and McCulloch (1993)) and the Gibb's sampler-based method of Geweke (1994). These algorithms all move randomly through the different models as a Markov chain approach and use results from the theory of Markov chain Monte Carlo's to derive theoretical convergence results. There are no analytic results concerning the relative computational efficiency of these algorithms.

In contrast we will take a simpler approach that matches the form of the prior distribution. We select models to evaluate by randomly including each variable with independent sampling probability $P_s(\beta_i)$. So long as the sampling probabilities are strictly greater than zero and strictly less than one, any values will work in the sense that, as the number of random draws grows with the sampled versions of (8), (9) and (10) will approach their true values.[16] Clyde, Desimone and Parmigiani (1996) have shown that this procedure, when implemented with $P_s(\beta_i)$ equal to the prior inclusion probability, (called by the authors "random sampling") has computational efficiency not importantly lower than that of the MC^3 and SVSS algorithms (for at least one particular data set). For the present application, we found that sampling models using their prior probabilities produced unacceptably slow convergence. Instead, we sampled one set of regressions using the prior probability sampling weights and then used the approximate posterior inclusion probabilities calculated from those regressions for the subsequent sampling probabilities. This results in "oversampling" well-fitting regressions and accelerates

---

[16] This is just the Law of Large Numbers at work.

convergence. Appendix 1 discusses computational and convergence issues in detail and may be of interest to researchers looking to apply these techniques.

3.- DATA

Hundreds of variables have been found to be significantly correlated with growth in the literature. Some of these variables are used systematically by most researchers. Others have been used only once. From all of these we selected 32 variables by using the following criteria.

First, we kept the variables that can, in some ways, represent *"state variables"* of a dynamic optimization problem. Hence, we choose variables measured as closely as possible to the beginning of the sample period (which is 1960) and eliminate all those variables that were computed for the later years only. For example, of all the education variables computed by Barro and Lee (1995), we only use the values for 1960. We also neglect some of the political variables that were published for the late 1980s, even though these variables have been put forward by a number of researchers (in this category, for example, we neglect Knack and Keefer's bureaucracy and corruption variables, which were computed for 1985 only; corruption and bad bureaucracy could very well be the endogenous response to a poor economic performance between 1960 and 1985).

Second, we also kept some variables, not because they are good proxies for some initial state variable but because they are proxies for "parameters" of some theoretical models, such as the rate of population growth for its role in the Solow model.

The third selection criterion derives from our need for a "balanced" data set. By balanced, we mean an equal number of observations for all regressions. Since different variables

19

miss observations for different countries, we selected the 32 variables that maximized the

product of the number of countries with observations for all variables and the number of

variables.

With these restrictions, the total size of the data set becomes 32 variables plus the growth

rate of GDP per capita between 1960 and 1992 for 98 countries. The variables, their means and

standard deviations are depicted in Table 1. Appendix 2 provides a list of the included countries.

4.- RESULTS

We are now ready to conduct our BACE estimation. We will calculate the posterior

distributions for all of the β's as well as the posterior means and variances given by equations (9)

to (10), using the posterior model weights from equation (8). We also calculate the posterior

inclusion probability, discussed in section 2.2, which provides a summary measure of how much

the data favor the inclusion of a particular variable in the growth regressions. Figure 1 shows the

posterior densities (approximated by histograms) of the coefficient estimates. Note that each

distribution consists of two parts: first, a continuous part that is the posterior density conditional

on inclusion in the model, and second, a discrete mass at zero representing the probability that

the variable does not belong in the model; this is given by one minus the posterior inclusion

probability[17]. As described in section 2, these densities are weighted averages of the posterior

densities conditional on each particular model with the weights given by the posterior model

---

[17] The probability mass at zero is split into ten bins around zero to make the area of the mass
comparable with areas under the conditional density. Also the maximum height of the lump at
zero is limited to 0.08 meaning that for some of the variables with very low inclusion probability
the rectangle shows slightly less probability mass than it actually has. All of the figures are
plotted with the same vertical axis scaling.

probabilities. A standard result from Bayesian theory (see, e.g., Leamer (1978) or Poirier (1995))

is that if prior are taken as diffuse by taking the limit of a Normal-Student prior[18] then the

posterior can be represented by:

$$t = \frac{\beta_i - \beta_{\hat{i}}}{s[(X'X)^{-1}]} \sim t(T-k) \tag{11}$$

where s is the usual OLS estimate of the standard deviation of the regression residual. In other

words, with the appropriate diffuse prior, the posterior distribution conditional on the model is

identical to the classical sampling distribution. Thus the marginal posterior distribution for each

coefficient is a *mixture-t* distribution. In principle these distributions could be of almost any

form, but most of the densities in Figure 1 look reasonably Gaussian.

4.1. Baseline Estimation

This section presents the baseline estimation results with a prior model size, $\bar{k}$, of seven

regressors. In Section 4.2 we examine the results with other values of $\bar{k}$. The results are based

on approximately 21 million randomly drawn regressions[19].

---

[18] That is a prior in which the marginal prior for the slope coefficients is multivariate normal and
the marginal prior for the regression error standard deviation is Student.

[19] The total number of possible regression models equals 2^{32}, which is approximately equal to 4.3
billion models. However, convergence of the estimates is attained relatively quickly; after 2
million draws the maximum change of coefficient estimates normalized by the standard deviation
of the regressors relative to the dependent variable is smaller than 10^{-5} per 10,000 draws, and
after 20 million draws the maximum change is smaller than 10^{-6}. The latter tolerance was used as
a convergence criterion for the reported estimates. See Appendix 1 for further details.

Table 2 shows the results: Column (1) reports the posterior inclusion probability of a variable in the growth regression. Columns (2) and (3) show the posterior mean and standard deviation of the distributions shown in figure 1, including the probability mass at zero. In contrast, columns (4) and (5) report the *conditional* posterior mean and standard deviation; that is, conditional on being included in the model. From the posterior density we can also calculate the posterior probability, conditional on inclusion, that a coefficient has the same sign as its posterior mean[20]; this "sign certainty probability" is contained in column (6). Finally, column (7) contains the (unweighted) fraction of regressions in which a coefficient is significantly different from zero in the classical sense of having a *t*-statistic with an absolute value greater than two.

In Table 2 the variables are sorted in descending order of their posterior inclusion probabilities. We can divide the variables according to whether seeing the data causes us to increase or decrease our inclusion probability relative to the prior probability: for the baseline estimation the prior inclusion probability is $7/32 = 0.219$. There are 12 variables for which the data provide support in this sense. For these variables, after seeing the data our belief that the variable belongs in the regression is strengthened. Among these 12 variables there is a natural division: (1) "top" variables that are strongly supported by the data with posterior inclusion probability above 0.95, (2) variables with some support by the data with inclusion probability below 0.95 but above the prior probability 0.219. The remaining 20 variables have little or no

---

[20] This "sign certainty probability" is analogous to the area under the normal CDF(0) calculated by Sala-i-Martin (1997 a,b).

support for inclusion: seeing the data further reduces our already modest initial assessment of their inclusion probability.

Recall that the posterior model probabilities are given by equation (8) with prior model probability given by the assumption that each variable has inclusion probability equal to 7/32. The posterior inclusion probability is the sum of the posterior probabilities of all of the regressions including that model. Thus, computationally, the posterior inclusion probability is a measure of the weighted average goodness of fit of models including a particular variable, relative to models not including the variable. The goodness of fit measure is adjusted to penalize highly parameterized models in the fashion of the Schwarz criterion. Thus, variables with high inclusion probabilities are variables which have high marginal contribution to the goodness-of-fit of the regression model. Readers uncomfortable with the Bayesian interpretation of the posterior inclusion probability may still regard this measure as a meaningful summary of the importance of a variable.

The posterior mean in column (2) is computed according to equation (9) while the posterior standard deviation is the square root of the variance formula in equation (10). The posterior mean is a weighted average of the OLS estimates for all regressions, including regressions in which the variable does not appear and thus has a coefficient of zero. The conditional mean in column (4) includes only the regressions in which the variable actually occurs and thus is equal to the posterior mean divided by the posterior inclusion probability. If one has the prior with which we began the estimation then the unconditional posterior mean is the "right" estimate of the marginal effect of the variable in the sense that it is the coefficient that

would be used for forecasting.[21] The conditional mean and variance are also of interest however.

From a Bayesian point of view these have the interpretation of the posterior mean and variance

for a researcher who has a prior inclusion probability equal to one for the particular variable

while maintaining the 7/32 inclusion probability for all the other variables. In other words, if one

is certain that the variable belongs in the regression, this is the estimate to consider. It is also

comparable to coefficient estimates in standard regressions not accounting for model uncertainty.

The conditional standard deviation provides one measure of how well estimated the variable is

conditional on its inclusion. It averages both the standard errors of each possible regression as

well as the dispersion of estimates across models.[22]

Examination of figure 1 and of the conditional means and standard deviations indicate

that all of the 12 variables which are supported by the data are also conditionally well estimated.

We will call these variables *robust*. While there may be combinations of conditioning variables

which lead to very different estimates for these coefficients, those regressions do not fit well and

receive low posterior weight. Further, on average these coefficients are well-estimated within

individual models. Note that there is in principle no reason why a variable could not have a very

---

[21] In a pure Bayesian approach there is not really a notion of a single estimate. However for
many purposes the posterior mean is reasonable, and it is what would be used for constructing
unbiased, minimum mean-squared-error predictions.

[22] Note that one cannot interpret the ratio of the posterior mean to the posterior standard deviation
as a *t*-statistic for two reasons. Firstly the posterior is a mixture *t*-distribution and secondly it is
not a sampling distribution. However, for most of the variables which we consider the posterior
distributions are not too far from being normal. To the extent to which these are approximately
normal, having a ratio of posterior conditional mean to standard deviation around two in absolute
value indicates an approximate 95% Bayesian coverage region that excludes zero.

high posterior inclusion probability and still be non-robust, it just happens that in our dataset there are no such variables.[23]

The "sign certainty probability" in column (6) is another measure of the significance of the variables. This is the posterior probability on the same side of zero as the posterior mean of the coefficient, conditional on the variable's inclusion. As noted above, for each individual regression the posterior density is equal to the classical sampling distribution of the coefficient. In classical terms, a coefficient would be 5% significant in a two-sided test if 97.5% of the probability in the sampling distribution were on the same side of zero as the coefficient estimate. So if, for example, it just happened that a coefficient were exactly 5% significant in every single regression its sign certainty probability would be 97.5%. Interestingly, applying a 0.975 cutoff to this quantity identifies exactly the same set of 12 variables as the examination of posterior probabilities (assuming we give the benefit of the doubt to the *real exchange rate distortions* variable at 0.974.) This reinforces our conclusion that these variables are robust.

The final column in table 2 reports the (unweighted) fraction of regressions in which the variable is classically significant at the 95% level. This is separated from the rest of the table because it was calculated separately from the other estimates.[24]   This is calculated partly for sake of comparison with extreme bounds analysis results. Note that for all but two of the

---

[23] This would occur if, for example, a variable contributed a great deal to the fit of the model but switched signs in the presence of another important variable.

[24] This column was calculated based on a run of 31 million regression. It was calculated separately so that the sampling could be based solely on the prior inclusion probabilities. The other baseline estimates were calculated by oversampling "good" variables for inclusion and thus produce misleading results for this statistic.

variables, many individual regressions can be found in which the variable is not significant, but even the two favored variables would still be labeled fragile by an extreme bounds test.

Another interesting statistic is the posterior mean model size. For this baseline estimation the prior model size was seven. But the data appear to favor somewhat larger models: the posterior mean model size is 9.9. This number is, of course, sensitive to the specification of the prior mean model size, as we will discuss below.

We are now ready to analyze the results from an economic point of view: what variables do we find to be "strongly" related to growth?

Variables Strongly and Robustly Related to Growth

The *Level of income in 1960* has an inclusion probability extremely close to one. The first panel in Figure 1 shows the posterior distribution of the coefficient estimates for initial income. Its inclusion probability is so high that in Figure 1, the mass at zero showing the probability that the variable does not belong in the model, is invisible. The posterior mean coefficient is -0.013 (with a standard deviation of 0.003); this is very precisely estimated. The implied speed of convergence[25] of 1.07% per year is somewhat smaller than 2% estimate given in Barro and Sala-i-Martin (1992). Due to the high inclusion probability the posterior mean is very close to the posterior mean conditional on inclusion. The sign certainty probability in column (6) shows that the probability mass of the density to the left of zero equals one to three decimal places: this can be seen in Figure 1 by the fact that almost all of the continuous density lies below

---

[25] The coefficient of convergence is given by $(1 - e^{-\beta T})/T$ (see Barro and Sala-i-Martin 1992).

zero. Notice that the fraction of regressions in which the coefficient for initial income has t-statistic greater than two in absolute value is only 59%, so that an extreme bounds test very easily labels the variable not robust. Nonetheless, the extremely well estimated coefficient and the very high sign certainty statistic show that initial income is indeed robustly partially correlated with growth. The explanation is that the regressions in which the coefficient on initial income is poorly estimated are regressions with very poor fit: thus they receive little weight in the averaging process. Furthermore, the extremely high inclusion probability suggests that regressions that omit initial income are likely to perform poorly.

The *Fraction of GDP in Mining* has a positive relationship with growth and a very high inclusion probability. This variable captures the success of countries with a large endowment of natural resources, although one might expect that large rents could also induce more political instability or rent-seeking.

The Sachs and Warner index of the *Number of Years an economy has been open between 1950 and 1994* has a strong positive association with growth. The openness index captures various aspects of the openness of a country to trade (tariff and non-tariff barriers, black market premium, degree of socialism and monopolization of exports by the government)[26].

---

[26] Rodriguez and Rodrik (1999) have recently criticized the Sachs-Warner index because it is mostly driven by the measure of the existence of state monopolies in major exports and by real exchange rate distortions whereas the other components add little explanatory power to the index. The Sachs-Warner index might therefore act as a catch-all variable for various macroeconomic difficulties and instability and Rodriguez and Rodrik warn to draw strong inferences about the effect of trade openness on growth from the coefficient.

The *fractions of Confucians* in the population enter positively in growth regressions. Note that the Confucian variable can be interpreted as a dummy for Hong Kong and some East Asian Tigers, which may explain the very high inclusion probability.

Variables Robustly Related to Growth

The following variables are supported by the data and are well-estimated. As will be discussed below however, the evaluation of their importance is more sensitive to prior beliefs than that of the variables above. Nonetheless, if one had a strong prior belief that any of these variables should belong in the model it would be judged to be statistically important.

Life expectancy in 1960 has a high inclusion probability and is positively related to subsequent growth of income per capita. Note that life expectancy may capture a whole set of factors (including nutrition, health care, social security, literacy rates) that are associated with high growth.

The *primary schooling enrolment rate in 1960* is positively related to growth and the inclusion probability is 0.63. Notice that when we consider larger prior model sizes in table 3, the inclusion probability of primary schooling increases dramatically (up to 0.95 with prior model size equal to 16) and falls for small prior model sizes. This suggests that primary schooling performs better in explaining growth when several other variables are included to capture steady state conditions.

Dummies for *Sub-Saharan Africa* and *Latin America* are negatively related to income growth. The posterior means are of comparable size, implying that Latin American and Sub-Saharan African countries had 0.6 and 0.7 percentage points lower income per capita growth

rates between 1960 and 1992, reducing the growth rate importantly from the sample average of 1.77.

The *fraction of Protestants* and growth are negatively related which may be explained by the relatively larger share of Protestants in OECD countries, which are closer to their steady states. For a country with 100% Protestants this effect is quite large at -0.6%.

The *fraction of primary exports in total exports* has a negative relationship with income growth with an economically large coefficient estimate. Given that almost all posterior weight is allocated to regressions containing the fraction of GDP in mining variable, the marginal effect of this variable is more focused on primary exports from non-mineral sources.

The *Real exchange rate distortions* variable is negatively related to income growth, but just barely make it into this category. Seeing the data only increases the inclusion probability by around 1% from the prior. The 11 variables ranked above this do much better. That said the variable is still quite robust and well-estimated with 97.4% of its conditional posterior density falling below zero.

Variables Marginally Related to Growth

The following four variables: *fraction of the population Buddhist, measure of outward orientation, war dummy* and the *index of political rights* all have posterior probabilities somewhat lower than their prior probabilities, but nonetheless are fairly robust if they are included in the growth regression. They all have sign certainty greater than 0.95, so that, very loosely speaking, they are on average 90% significant. As we will discuss below the exact importance of these variables is somewhat subject to the specific prior beliefs.   The *measure of*

outward orientation has a somewhat surprising negative partial correlation with growth. Note that the very high inclusion probability for the *years of openness* variable means that there is already one openness measure in the regression.

Variables Not Robustly Related to Growth

The remaining sixteen variables show little evidence of any sort of robust partial correlation with growth. They neither contribute importantly to the goodness-of-fit of growth regressions, as measured by their posterior inclusion probabilities, nor have estimates that are robust across different sets of conditioning variables. Notice that some political variables such as the *number of revolutions and coups* or the *index of political rights* are not robustly related to economic growth. Similarly the *degree of capitalism* measure has no positive correlation with growth. This could be due to the fact that other variables which capture political or economic instability such as real exchange rate distortions, the number of years an economy has been open and life expectancy or regional dummies capture most of the variation in those variables.

Overall, we can conclude that the data support the hypothesis that there is a set of variables that are robustly partially correlated with economic growth. This strongly contradicts the extreme bounds test which would reject the significance of every single variable. Further note that the list of robust variables is similar to that reported by Sala-i-Martin (1997), but that several of the variables that Sala-i-Martin identified as significantly related to growth have low inclusion probabilities. This can be reconciled by the fact that the weighting method used by Sala-i-Martin (1997), weighting by the model likelihood, assigns a lot of weight to a very few

regressions. By contrast, the weighting method derived in this paper puts less extreme emphasis on small groups of regressions and incorporates the estimates from a much larger set of models.

4.2. Robustness of Results

Up until now we have concentrated on results derived for a prior model size $\bar{k} = 7$. As discussed earlier, while we feel that this is a reasonable expected model size it is in some sense arbitrary. We need to explore the effects of the prior on our conclusions. Tables 3, 4 and 5 do precisely this, reporting the posterior inclusion probabilities and conditional posterior means, respectively, for $\bar{k}$ equal to 5, 9, 11 and 16 as well as repeating the benchmark numbers for easy comparison.

First note that the results for the four strongest variables show almost no sensitivity whatsoever to the choice of prior model size, either in terms of their inclusion probabilities or their coefficient estimates. On the other end of the scale, the sixteen variables that showed little partial correlation in the baseline estimation are not helped by alternative priors. Their posterior inclusion probabilities rise as $\bar{k}$ increases, which is hardly surprising as their prior inclusion probabilities are rising. But even the best in this group always have posterior inclusion probabilities far lower than their prior inclusion probabilities. For example, one of the best in a poor lot is the *civil liberties index* (civlibb) which has a posterior probability of around 25% when $\bar{k} = 16$. But with $\bar{k} = 16$ the prior inclusion probability is 50% so that the data halve our prior probability.   For some of the variables in this group the coefficient estimates are not very stable across different model specifications but since they are not robust in the first place this is of little importance.

More interesting are the in-between variables, some of which display interesting systematic patterns when $\bar{k}$ is varied. One of these is *primary school enrolment rate in the initial period (P60)*: here the posterior inclusion probability rises from 49% with $\bar{k} = 5$ to 95% with $\bar{k} = 16$. This suggests that the primary school enrolment rate is a variable which requires other conditioning variables in order to display its full importance. Interestingly, the conditional estimate of its slope coefficient is quite stable across specifications. Both the *fraction protestant* and the *primary commodity export share* are also variables which appear to do better with more conditioning variables and also have stable coefficient estimates. Unsurprisingly, the coefficients on the two regional dummies (Latin America and Sub-Saharan Africa) decline in absolute value as more conditioning variables are included. Only one variable in the list of 12 that are robust in the baseline regressions would drop out when priors indicate larger model sizes: real exchange rate distortions (RERD). For $\bar{k} \geq 9$ the posterior probability is lower than the prior and the coefficient is falling. This suggests that this variable is acting more as a catchall for various other effects than in its own right. All of the other 11 variables that were robust in the baseline model also appear to be robust to different prior specifications as can be seen by examining the sign certainty probabilities in Table 5.

Finally there are the four marginally important variables from the baseline. Of these, two, *fraction Buddhist* and the *political rights index*, show little change: there is still marginal evidence for their inclusion. The other two, the *index of outward orientation*, and the *war dummy* have posterior inclusion probabilities strongly increasing with the prior model size. For both, in the somewhat extreme $\bar{k} = 16$ case the posterior probability is above the prior inclusion probability.

32

Unsurprisingly the posterior mean model size is strongly affected by the choice of prior mean model size. Recall that for the baseline $\bar{k} = 7$ the posterior model size was 9.9. For prior model sizes of five through 11 the posterior model size is greater than the prior, indicating that the data prefer larger models. With a prior model size of 16 the data reduce our opinion of the size of the model. The posterior model size is quite sensitive to the prior specification.

The overall conclusions are at most mildly affected by considering different prior model sizes. Including robustness to different prior specifications we still find 11 variables importantly and robustly correlated with growth and five variables with marginal correlation.

5.- CONCLUSIONS

In this paper we propose a Bayesian Averaging of Classical Estimates method to determine what variables are strongly related to growth in a broad cross section of countries. The method introduces a number of improvements relative to the previous literature. For example, we improve upon Sala-i-Martin (1997) because we use an averaging method which is a (approximately) fully justified Bayesian estimators and because we do not restrict the number regressors in the averaged models. Our approach provides an alternative to a standard Bayesian Model Averaging since BACE does not require the specification of the prior distribution of the parameters, but has only one hyper-parameter, the expected model size, $\bar{k}$. This parameter is easy to interpret, easy to specify, and easy to check for robustness. The interpretation of the BACE estimates is straightforward for economists not trained in Bayesian inference, since the weights are analogous to the Schwarz model selection criterion. Finally, our estimates can be calculated using only repeated applications of OLS which makes the approach transparent and

33

straightforward to implement. In contrast to extreme bounds tests, models that fit poorly are not given equal weight with those that fit well and no variables are held "fixed" and therefore "untested."

Our main results support Sala-i-Martin rather than Levine and Renelt: we find that a good number of economic variables have robust partial correlation with long-run growth. In fact, we find that about one third of the 32 variables used in the analysis can be said to be robustly related to growth while several more are marginally related. Interestingly, the strongest variable is the initial level of income which reflects the concept of conditional convergence discussed in Barro and Sala-i-Martin (1992). Other important variables include regional dummies (such as Africa or Latin America), some measures of human capital (such as life expectancy or primary schooling) and some sectoral variables such as measures of openness, primary exports or real exchange distortions.

There are three lines of research that we would like to pursue from here. First we would like to apply a version of our BACE method to panel data estimates, which have become quite popular in the recent empirical economic growth literature. Secondly, we plan to allow the inclusion of nonlinear terms in the regressions. The literature has identified some variables which may affect growth in a highly nonlinear way: for example, it has been argued that inflation has important negative effects on growth, but only for very high levels of inflation. Our analysis forces all variables to enter the regressions in a linear fashion and, therefore, does not allow for such nonlinearities. Finally, we would like to expand our theoretical and empirical analysis to allow for unbalanced data sets. The lack of data is an important problem for many fields, but it is an especially important problem for the field of cross-country comparisons of long-term data like

the one required by the growth literature. Our analysis so far has required that the number of

observations be identical for each of the regressions and, as a result, we have been forced to

neglect many of the potentially important variables simply because they were missing too many

observations.

APPENDIX 1

This appendix includes some more precise details about computational aspects of the BACE procedure with particular emphasis on the sampling algorithm and convergence. Given the form of our prior distribution, the prior inclusion probability for each variable is $\bar{k}/K$ as described in the main text. Represent a model, M_j, as a length K binary vector in which a one indicates that a variable is included in the model and a zero indicates that it is not. Then:

$$P(M_j) = \left[\prod_{i=1}^{K} M_{ji} \frac{\bar{k}}{K} \right] \cdot \left[\prod_{i=1}^{K} (1 - M_{ji})(1 - \frac{\bar{k}}{K}) \right] = \left(\frac{\bar{k}}{K} \right)^{k_j} \cdot \left(1 - \frac{\bar{k}}{K} \right)^{1-k_j} \qquad (A1)$$

where k_j is the number of included variables in model j and M_{ji} is the i'th element of the M_j vector. The second equality in (A1) holds only in the case of equal prior inclusion probabilities for each variable, but the first equality is easily adapted to the case in which the prior inclusion probabilities may differ across variables. If the set of possible regressions is small enough to allow exhaustive calculation, one may substitute (A1) into (8) to calculate the posterior model probabilities and then use (9) and (10) to calculate the posterior mean and variance. For each term of the sum one calculates the appropriate OLS regression, gets the OLS parameter estimates for the β's and σ and the sum of squared errors. These allow the computation of the individual term in (9) and (10). Also the posterior probabilities allow the calculation of any other features of the posterior distribution which may be of interest based on the 2^K-term version of (4). As for the other quantities cited in this paper, the "sign-certainty statistic" is given by:

$$sign\ certainty\ for\ \beta_j = \sum_{j=1}^{2^K} P(M_j|y) P\left[sign(\beta_j) = sign\left(E(\beta_j|y)\right)|M_j, y\right] \qquad (A2)$$

The histograms for the posterior densities are calculated as follows. An initial run established the important range of the distribution of the estimates for each β. This was then split into 100 equal size bins for the histogram. Since for each regression the ratio of $\hat{\beta}$ to the estimated standard deviation of the error term is distributed t(T-k-1) we can use a t-CDF to evaluate the amount of probability contained in each bin. This is then weighted by the posterior probability of the regression. Note that the calculation of these histograms is quite computationally intensive as with each regression we must make 100 times k calls to a *t*-CDF.

When we are sampling randomly from the space of possible models we want the limits of all of our quantities of interest to approach their true values as the number of sampled models approaches infinity. If we let the probability of sampling M_j be given by $P_s(M_j)$ then the weight attached to each regression must be adjusted by the inverse of the sampling probability. This is because as the number of sampled regressions approaches infinity the fraction of times a particular regression is run approaches its sampling probability, when in sums such as (9) and (10) each regression gets equal weight. Thus, with sampling the analog of (8) becomes:

$$P(M_j|y) = \frac{\dfrac{P(M_j)}{P_s(M_j)} T^{-k_j/2} SSE_j^{-T/2}}{\left(\displaystyle\sum_{i=1}^{N} \dfrac{P(M_{m(i)})}{P_s(M_{m(i)})} T^{-k_{m(i)}/2} SSE_{m(i)}^{-T/2} \right)} \qquad \text{(A3)}$$

where $m(i)$ represents the model index associated with the i'th randomly sampled model and N is the number of models sampled. This version of the weights can then be used to calculate sampling analogs of (9) and (10). The intuition for (A3) is that we are over-sampling some

models so as usual we have to deflate observations by their sampling probabilities. (A3) is particularly easy to calculate when the sampling probabilities are equal to the prior probabilities in which case they cancel and need not even be computed. This is the sampling strategy discussed in the text of randomly selecting models by randomly including variables with their initial inclusion probability. So long as the sampling probabilities of all models are greater than zero all of the numerical approximations will be consistent.

Trial-and-error calculation indicated that for the present problem the prior-weight sampling was leading to slow convergence of the parameter estimates. This is because it samples many, many poorly fitting regressions which receive little weight in the averages. Instead we used the following procedure which we refer to as the stratified sampler: we ran 100,000 regressions using the prior weight sampler and then adjusted the sampling inclusion probabilities to be equal to the posterior inclusion probabilities estimated from the initial sample. In order to guard against too much impact from errors made in the first 100,000 regressions we limited the sampling probabilities to lie in the interval [0.1, 0.85]. Some experimentation suggested that moderate changes in these bounds has little effect on the behavior of the algorithm. Again, since any set of sample inclusion probabilities will work asymptotically the choice of these parameters is not critical. Thus our stratified sampler over-samples "good" regressions.

We then need some way of judging whether or not the sampled analogs of (9) and (10) are approaching their limits. As always, convergence criteria are somewhat arbitrary. For the estimates reported in the paper we examined changes in the posterior means of the β's. First we normalized the coefficient estimates by the ratio of the standard deviation of y to the standard deviation of x. The standardization with respect to y is only to make the size of the convergence

criterion easy to interpret. This transformation standardizes the β's into units of standard deviations of y per standard deviation of x. Then in order to declare that the estimates "converged" we looked at the change in the estimates of the normalized β's generated by sampling a further 10,000 regressions. When this change fell below 10E-06 for ten consecutive sets of 10,000 regressions the algorithm declared convergence. For our stratified sampling technique these parameter changes fall smoothly as a function of the number of regressions so that this criterion is reasonable. For the prior probability sampler this change is much less reliable with the occasional set of 10,000 having a large impact: we would not recommend the use of this sampler with this particular convergence criterion. For our baseline estimation with $\overline{K} = 7$ we also investigated the performance of the sampler and convergence criterion by performing a number of further runs with the same convergence criterion: these all converged with between 15 and 30 million regressions. Results were very similar: they suggest that the posterior inclusion probabilities in table 2 are accurate to at least two decimal places, while the conditional β estimates are even more accurate. Estimates based on only two million or so regressions are even quite close to the 20 million regression baseline. This suggests that our methodology will create quite accurate approximations in reasonable computing times even with very large model spaces.

In the Bayesian Model Averaging statistics literature, which has used fully Bayesian estimates of individual models, the most popular sampling algorithm appears to be the MC^3 algorithm mentioned in the main text. We were resistant to using this algorithm because its mechanism, based on the Metropolis-Hastings criterion, is quite difficult to understand intuitively. In order, however, to both try to ensure that our stratified sampler is generating

correct answers and to compare it to procedures in other work we created a test data set. This used all of the observations in our main data set but with only 20 variables rather than the full set of 32. This reduces the set of possible regressions to around one million which easily allows the precise calculation of the sums in (8), (9) and (10). We then performed sampling runs with 50,000 regressions each and calculated a weighted mean-squared error criterion for the posterior means of the β's with the weighting matrix being $(X'X)^{-1}$. By this criterion the stratified sampling algorithm was about four times as accurate as MC3, but for both accuracy was quite reasonable.

APPENDIX 2

Countries included in the regressions:

Algeria	El Salvador	Cyprus
Benin	Guatemala	Denmark
Botswana	Haiti	Finland
Burundi	Honduras	France
Cameroon	Jamaica	Germany, West
Cent'l Afr. Rep.	Mexico	Greece
Chad	Nicaragua	Ireland
Congo	Panama	Italy
Egypt	Trinidad & Tobago	Netherlands
Ethiopia	United States	Norway
Gabon	Argentina	Portugal
Gambia	Bolivia	Spain
Ghana	Brazil	Sweden
Guinea-Bissau	Chile	Switzerland
Kenya	Colombia	Turkey
Lesotho	Ecuador	United Kingdom
Liberia	Guyana	Australia
Madagascar	Paraguay	Fiji
Malawi	Peru	New Zealand
Mali	Uruguay	Papua New Guinea
Mauritania	Venezuela	
Mauritius	Hong Kong	
Morocco	India	
Niger	Indonesia	
Nigeria	Israel	
Rwanda	Japan	
Senegal	Jordan	
Somalia	Korea	
South africa	Malaysia	
Tanzania	Nepal	
Togo	Pakistan	
Tunisia	Philippines	
Uganda	Singapore	
Zaire	Sri Lanka	
Zambia	Syria	
Zimbabwe	Taiwan	
Canada	Thailand	
Costa Rica	Austria	
Dominican Rep.	Belgium	

REFERENCES

Barro, Robert J. (1991a). "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, **106**, 2 (May), 407-443.

Barro, Robert, (1996) "Determinants of Democracy", Mimeo Harvard University, July.

Barro, Robert J. and Jong-Wha Lee (1993). "International Comparisons of Educational Attainment," *Journal of Monetary Economics*, 32, 3 (December), 363-394. The data for this paper were taken from the NBER Web Page.

Barro, Robert J and Xavier Sala-i-Martin (1992) "Convergence," *Journal of Political Economy,* **100**(2), 223-51.

Barro, Robert J and Xavier Sala-i-Martin (1995), *Economic Growth,* McGraw Hill.

Caselli, Francesco; Gerard Esquivel and Fernando Lefort (1996), "Reopening the convergence debate: a new look at cross-country growth empirics," *Journal of Economic Growth* **1**(3): 363-89.

Clyde, M.; Desimone, H., and Parmigiani, G. (1996), "Prediction via Orthogonalized Model Mixing," *Journal of the American Statistical Association*, **91**, 1197-1208.

DeLong and Summers, (1991) "Equipment Investment and Economic Growth", *Quarterly Journal of Economics,* 106, 2 (May), 445-502. The data for this paper was taken from the World Bank's Research Department Web Page.

Doppelhofer, Gernot (2000), *Doctoral Dissertation*, Columbia University.

Durlauf, Steven N. and Danny T. Quah (1999), "The New Empirics of Economic Growth," in *Handbook of Macroeconomics Vol. 1*, John B. Taylor and M. Woodford, eds., North Holland, Amsterdam.

Easterly William and Ross Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics,* 112, 4 (November), 1203-50.

Fernandez, Carmen; Ley, Eduardo and Mark F. J. Steel (2000), "Model Uncertainty in Cross-Country Growth Regressions," mimeo.

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

Geweke, J.F. (1994), "Bayesian Comparison of Econometric Models," Working Paper 532, Federal Reserve Bank of Minneapolis.

Granger, Clive W. J. and Harald Uhlig (1990), "Reasonable Extreme-Bounds Analysis," *Journal of Econometrics*, **44**, 159-170.

Hoeting, Jennifer; Madigan, David; Raftery, Adrian, and Chris Volinsky (1999), "Bayesian Model Averaging: A Tutorial," Technical Report 9814, Department of Statistics, Colorado State University.

Islam, Nazrul, (1995) "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, **110**(4), 1127-70.

Jeffreys (1961) *Theory of Probability*, 3rd ed., Oxford University Press, London.

Hall, Robert and Charles Jones, (1996), "The Productivity of Nations", NBER Working Paper #5812, November 1996. The data for this paper were taken from the Chad Jone's Web Page.

King, Robert G., and Ross Levine. 1993. "Finance, Entrepreneurship, and Growth: Theory and Evidence." *Journal of Monetary Economics* **32** (3): 513-42.

Knack, Stephen and Philip Keefer (1995), "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures", *Economics and Politics*. The data from this paper were provided to us by Robert Barro.

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

Leamer, Edward E. (1983). "Let's take the con out of econometrics", *American Economic Review*, 73, 1, (March), 31-43.

Leamer, Edward E. (1985). "Sensitivity Analysis Would Help", *American Economic Review*, 75, 3, (June), 308-313.

Levine, Ross and David Renelt (1992). "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, **82**, 4 (September), 942-963. The data for this paper was taken from the World Bank's Research Department Web Page.

Madigan, D., and York, J. (1995), "Bayesian Graphical Models for Discrete Data," *International Statistical Review*, **89**, 215-232.

Raftery, Adrian; Madigan, David; and Jennifer Hoeting (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, **92**(437), 179-191.

Rodriguez, Francisco and Dani Rodrik (1999), "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence", *NBER Working Paper # 7081*.

Sachs, Jeffrey and Andrew Warner, (1995), "Economic Reform and the Process of Economic Integration", *Brookings Papers of Economic Activity,* #1, (August), pp1-95. The data for this paper was provided to me by Andrew Warner.

Sachs, Jeffrey and Andrew Warner, (1996) "Natural Resource Abundance and Economic Growth", mimeo HIID.

Sala-i-Martin, X., (1997a), "I Just Ran 2 Million Regressions", *American Economic Review*, May.

Sala-i-Martin, X., (1997b), "I Just Ran Four Million Regressions", *NBER Working Paper #6252*.

Schwarz, G. (1978) "Estimating the Dimension of a Model," *The Annals of Statistics*, **6**, 461-464.

York, J, Madigan, D., Heuch, I. and R. T. Lie (1995), "Estimating a Proportion of Birth Defects by Double Sampling: A Bayesian Approach Incorporating Covariates and Model Uncertainty," *Applied Statistics*, **44**, 227-242.

Zellner, Arnold (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zellner, Arnold (1986), "On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions," in *Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti. Studies in Bayesian Econometrics and Statistics series, vol. 6,* Goel, Prem, and Zellner, Arnold eds., North-Holland, Amsterdam, 233-43.

Table 1: Description of Data and Sources of Variables

	Name	Variable Description and Source	Mean	S.D.
	Growth	Growth of GDP per capita between 1960 and 1992. Barro and Lee (1993) [henceforth BL93]	0.0177	0.0180
1	GDPSH60	log(GDP per capita 1960). Log of Summers-Heston GDP per capita in 1960. BL93.	7.3273	0.9046
2	LIFEE060	Life Expectancy in 1960. BL93.	53.4173	12.2979
3	P60	Primary School Enrollment Rate in 1960. BL93.	0.7143	0.3064
4	safrica	Sub-Sahara African Dummy. Dummy for Sub-Sahara African Countries.	0.3265	0.4714
5	laam	Latin American Dummy. Dummy for Latin American countries.	0.2245	0.4194
6	OECD	OECD Dummy. Dummy for OECD countries.	0.2245	0.4194
7	SCOUT	Outward Orientation. Dummy for outward orientation. Levine and Renelt (1992).	0.3673	0.4846
8	dpop6090	Growth Rate of Population between 1960 and 1990. BL93.	0.0213	0.0094
9	h60	Higher Education Enrollment Rate in 1960. BL93.	0.0348	0.0486
10	YrsOpen	Number of Years economy has been Open between 1950 and 1994. Index computed by Sachs and Warner (1995).	0.3616	0.3504
11	revcoup	Revolutions and Coups. Number of military coups and revolutions. BL93.	0.1844	0.2290
12	wardum	War Dummy. Dummy for countries that have been involved in war any time between 1960 and 1990. BL93	0.3878	0.4897
13	prightsb	Political Rights. See Barro (1996). Larger index values indicate fewer rights. [henceforth B96]	3.8801	2.0197
14	civlibb	Civil Liberties. Index of civil liberties. Knack and Keefer (1995). Larger values indicate fewer civil liberties.	3.8362	1.8183
15	ABSLATIT	Absolute Latitude. B96.	22.7228	16.4984
16	AVELF	Index of Ethnolinguistic Fractionalization. Probability two random people in a country do not speak same language. From Easterly and Levine (1997).	0.3617	0.3037
17	PRIEXP70	Primary Exports in 1970. Fraction of primary exports in total exports in 1970. Sachs and Warner (1996)	0.7330	0.2826
18	RERD	Real Exchange Rate Distortions. Levine and Renelt (1992).	125.4694	40.7523
19	BRIT	British Colony. Dummy variable for former British colonies. See B96.	0.3367	0.4750
20	FRENCH	French Colony. Dummy variable for former French colonies. See B96.	0.1837	0.3892
21	SPAIN	Spanish Colony. Dummy variable for former Spanish colonies. See B96.	0.1633	0.3715
22	BUDDHA	Fraction of Buddhists. See B96.	0.0418	0.1594
23	CATH	Fraction of Catholics. See B96.	0.3554	0.3691
24	CONFUC	Fraction of Confucians. Fraction of population that follows Confucian Religion. See B96.	0.0140	0.0753
25	HINDU	Fraction Hindu. See B96.	0.0379	0.1437
26	JEW	Fraction Jewish. See B96.	0.0093	0.0828
27	MUSLIM	Fraction Muslim. See B96.	0.2033	0.3345
28	PROT	Fraction Protestant. See B96.	0.1688	0.2312
29	Mining	Fraction of GDP in Mining. From Hall and Jones (1996).	0.0479	0.0735
30	EcOrg	Degree of Capitalism. Index of degree in which economies favor capitalist forms of production. Hall and Jones (1996).	3.4388	1.4436
31	OthFrac	Fraction of Population Speaking Foreign Language. Hall and Jones (1996).	0.3123	0.4051
32	EngFrac	Fraction of Population Speaking English. Fraction of the population able to speak English. Hall and Jones (1996).	0.0841	0.2517

Table 2: Baseline Estimation

	Posterior Inclusion Probability	Posterior Mean	Posterior Standard Deviation	Posterior Mean Conditional on Inclusion	Conditional Posterior Standard Deviation	"Sign Certainty Probability"	Fraction of Regressions with Abs(t)>2
1 GDPSH60	1.000	-0.01276	0.00274	-0.01276	0.00273	1.000	0.59
29 Mining	0.998	0.06475	0.01490	0.06490	0.01458	1.000	0.54
10 YrsOpen	0.997	0.01819	0.00447	0.01824	0.00436	1.000	1.00
24 CONFUC	0.971	0.05808	0.01861	0.05983	0.01588	1.000	1.00
2 LIFEE060	0.887	0.00079	0.00039	0.00088	0.00028	0.999	0.71
3 P60	0.627	0.01214	0.01107	0.01938	0.00744	0.995	0.89
4 safrica	0.596	-0.00720	0.00682	-0.01209	0.00436	0.994	0.85
27 MUSLIM	0.580	0.00809	0.00779	0.01395	0.00480	0.994	0.26
5 laam	0.514	-0.00588	0.00642	-0.01143	0.00409	0.992	0.62
28 PROT	0.474	-0.00641	0.00781	-0.01351	0.00571	0.991	0.71
17 PRIEXP70	0.430	-0.00592	0.00776	-0.01377	0.00564	0.989	0.76
18 RERD	0.239	-0.00002	0.00003	-0.00007	0.00003	0.974	0.52
22 BUDDHA	0.194	0.00271	0.00638	0.01398	0.00724	0.967	0.96
7 SCOUT	0.151	-0.00059	0.00164	-0.00392	0.00218	0.956	0.01
12 wardum	0.145	-0.00058	0.00164	-0.00396	0.00227	0.955	0.23
13 prightsb	0.138	-0.00023	0.00068	-0.00165	0.00099	0.950	0.40
32 EngFrac	0.107	-0.00077	0.00268	-0.00719	0.00464	0.933	0.35
14 civlibb	0.104	-0.00017	0.00064	-0.00164	0.00123	0.917	0.26
23 CATH	0.096	-0.00073	0.00304	-0.00759	0.00660	0.855	0.45
11 revcoup	0.078	-0.00052	0.00225	-0.00663	0.00491	0.905	0.08
31 OthFrac	0.076	0.00032	0.00149	0.00429	0.00352	0.882	0.30
15 ABSLATIT	0.073	0.00001	0.00005	0.00015	0.00013	0.874	0.40
30 EcOrg	0.054	0.00004	0.00025	0.00081	0.00075	0.855	0.15
16 AVELF	0.054	-0.00026	0.00156	-0.00486	0.00478	0.843	0.30
8 dpop6090	0.045	0.00719	0.05220	0.15920	0.19000	0.796	0.15
6 OECD	0.044	0.00014	0.00135	0.00324	0.00557	0.716	0.21
25 HINDU	0.044	-0.00016	0.00212	-0.00361	0.00943	0.644	0.02
21 SPAIN	0.044	-0.00009	0.00123	-0.00199	0.00551	0.647	0.36
19 BRIT	0.034	0.00001	0.00048	0.00033	0.00257	0.544	0.12
26 JEW	0.033	0.00020	0.00256	0.00594	0.01277	0.677	0.00
20 FRENCH	0.032	0.00000	0.00058	0.00007	0.00323	0.502	0.20
9 h60	0.031	0.00001	0.00559	0.00047	0.03176	0.503	0.06

Notes: The left hand side variable in all regressions is the growth rate from 1960-1992 across 98 countries. Apart from the final column all statistics come from a random sample of approximately 21 million of the possible regressions including any combination of the 32 variables. Prior mean model size is seven. Variables are ranked by the first column, the posterior inclusion probability. This is the sum of the posterior probabilities of all models containing the variable. The next two columns reflect the posterior mean and standard deviations for the linear marginal effect of the variable: the posterior mean has the usual interpretation of a regression β. The conditional mean and standard deviation are conditional on inclusion in the model. The "sign certainty probability" is the posterior probability that the coefficient is on the same side of zero as its mean conditional on inclusion. It is a measure of our posterior confidence in the sign of the coefficient. The final column is the fraction of regressions in which the coefficient has a classical t-test greater than two, with all regressions having equal sampling probability.

Table 3: Posterior Inclusion Probabilities with Different Prior Model Sizes $\bar{k}$

	$\bar{k} = 5$	$\bar{k} = 7$	$\bar{k} = 9$	$\bar{k} = 11$	$\bar{k} = 16$
Prior Inclusion Probability	0.156	0.219	0.281	0.344	0.500
1 GDPSH60	0.999	1.000	1.000	1.000	1.000
29 Mining	0.996	0.998	0.999	0.999	1.000
10 YrsOpen	0.996	0.997	0.997	0.997	0.997
24 CONFUC	0.974	0.971	0.970	0.971	0.977
2 LIFEE060	0.898	0.887	0.880	0.875	0.868
3 P60	0.488	0.627	0.740	0.825	0.947
4 safrica	0.589	0.596	0.606	0.624	0.715
27 MUSLIM	0.534	0.580	0.630	0.679	0.774
5 laam	0.511	0.514	0.526	0.552	0.668
28 PROT	0.390	0.474	0.544	0.605	0.740
17 PRIEXP70	0.318	0.430	0.526	0.615	0.788
18 RERD	0.223	0.239	0.250	0.256	0.271
22 BUDDHA	0.137	0.194	0.254	0.316	0.469
7 SCOUT	0.086	0.151	0.235	0.331	0.596
12 wardum	0.084	0.145	0.220	0.307	0.555
13 prightsb	0.103	0.138	0.178	0.212	0.293
32 EngFrac	0.084	0.107	0.127	0.146	0.198
14 civlibb	0.069	0.104	0.139	0.175	0.257
23 CATH	0.065	0.096	0.129	0.162	0.248
11 revcoup	0.056	0.078	0.099	0.119	0.167
31 OthFrac	0.064	0.076	0.088	0.101	0.143
15 ABSLATIT	0.054	0.073	0.093	0.108	0.149
30 EcOrg	0.036	0.054	0.075	0.099	0.183
16 AVELF	0.038	0.054	0.073	0.092	0.149
8 dpop6090	0.031	0.045	0.063	0.084	0.159
6 OECD	0.030	0.044	0.061	0.078	0.131
25 HINDU	0.027	0.044	0.064	0.088	0.178
21 SPAIN	0.032	0.044	0.057	0.073	0.136
19 BRIT	0.023	0.034	0.047	0.063	0.116
26 JEW	0.023	0.033	0.046	0.060	0.107
20 FRENCH	0.021	0.032	0.044	0.060	0.116
9 h60	0.020	0.031	0.043	0.056	0.102

Notes: The left hand side variable in all regressions is the growth rate from 1960-1992 across 98 countries. Each column contains the posterior probability of all models including the given variable. These are calculated with the same data but which different prior mean model sizes as labeled in the column headings. They are based on different random samples of all possible regressions using the same convergence criterion for stopping sampling. Samples range from around 4 million regressions for $\bar{k} = 5$ to around 30 million for $\bar{k} = 16$.

Table 4: Posterior Conditional Means with Different Prior Model Sizes $\bar{k}$

		$\bar{k} = 5$	$\bar{k} = 7$	$\bar{k} = 9$	$\bar{k} = 11$	$\bar{k} = 16$
	Prior Inclusion Probability	Posterior Mean Conditional on Inclusion				
1	GDPSH60	-0.01274	-0.01276	-0.01281	-0.01285	-0.01291
29	Mining	0.06502	0.06490	0.06479	0.06481	0.06513
10	YrsOpen	0.01876	0.01824	0.01776	0.01728	0.01604
24	CONFUC	0.06153	0.05983	0.05846	0.05718	0.05437
2	LIFEE060	0.00095	0.00088	0.00083	0.00078	0.00068
3	P60	0.01933	0.01938	0.01958	0.01990	0.02080
4	safrica	-0.01281	-0.01209	-0.01142	-0.01083	-0.00980
27	MUSLIM	0.01409	0.01395	0.01372	0.01344	0.01264
5	laam	-0.01176	-0.01143	-0.01097	-0.01046	-0.00957
28	PROT	-0.01354	-0.01351	-0.01344	-0.01338	-0.01325
17	PRIEXP70	-0.01383	-0.01377	-0.01370	-0.01361	-0.01341
18	RERD	-0.00007	-0.00007	-0.00006	-0.00006	-0.00005
22	BUDDHA	0.01412	0.01398	0.01392	0.01384	0.01350
7	SCOUT	-0.00372	-0.00392	-0.00408	-0.00421	-0.00449
12	wardum	-0.00380	-0.00396	-0.00406	-0.00416	-0.00438
13	prightsb	-0.00171	-0.00165	-0.00160	-0.00155	-0.00143
32	EngFrac	-0.00761	-0.00719	-0.00678	-0.00636	-0.00539
14	civlibb	-0.00165	-0.00164	-0.00162	-0.00159	-0.00146
23	CATH	-0.00784	-0.00759	-0.00724	-0.00686	-0.00603
11	revcoup	-0.00688	-0.00663	-0.00634	-0.00599	-0.00495
31	OthFrac	0.00477	0.00429	0.00383	0.00339	0.00257
15	ABSLATIT	0.00016	0.00015	0.00014	0.00013	0.00009
30	EcOrg	0.00082	0.00081	0.00081	0.00080	0.00080
16	AVELF	-0.00504	-0.00486	-0.00467	-0.00449	-0.00399
8	dpop6090	0.16213	0.15920	0.15926	0.16122	0.17161
6	OECD	0.00300	0.00324	0.00330	0.00318	0.00267
25	HINDU	-0.00200	-0.00361	-0.00452	-0.00558	-0.00768
21	SPAIN	-0.00293	-0.00199	-0.00117	-0.00035	0.00162
19	BRIT	0.00028	0.00033	0.00037	0.00043	0.00071
26	JEW	0.00663	0.00594	0.00541	0.00483	0.00306
20	FRENCH	0.00046	0.00007	-0.00024	-0.00062	-0.00134
9	h60	-0.00229	0.00047	0.00198	0.00365	0.00605

Notes: The left hand side variable in all regressions is the growth rate from 1960-1992 across 98 countries. Each column contains the posterior mean of the regression slope coefficient for the given variable conditional on the variables inclusion in the model. These are calculated with the same data but which different prior mean model sizes as labeled in the column headings. They are based on different random samples of all possible regressions using the same convergence criterion for stopping sampling. Samples range from around 4 million regressions for $\bar{k} = 5$ to around 30 million for $\bar{k} = 16$.

Table 5: Sign Certainty Probabilities
with Different Prior Model Sizes $\bar{k}$

	$\bar{k}=5$	$\bar{k}=7$	$\bar{k}=9$	$\bar{k}=11$	$\bar{k}=16$
1 GDPSH60	1.000	1.000	1.000	1.000	1.000
29 Mining	1.000	1.000	1.000	1.000	1.000
10 YrsOpen	1.000	1.000	1.000	1.000	0.999
24 CONFUC	1.000	1.000	0.999	0.999	0.999
2 LIFEE060	1.000	0.999	0.998	0.997	0.994
3 P60	0.992	0.995	0.996	0.995	0.997
4 safrica	0.999	0.994	0.987	0.991	0.986
27 MUSLIM	0.996	0.994	0.992	0.993	0.990
5 laam	0.996	0.992	0.988	0.986	0.983
28 PROT	0.993	0.991	0.986	0.989	0.986
17 PRIEXP70	0.992	0.989	0.990	0.990	0.990
18 RERD	0.966	0.974	0.962	0.958	0.922
22 BUDDHA	0.973	0.967	0.966	0.965	0.965
7 SCOUT	0.925	0.956	0.952	0.964	0.976
12 wardum	0.962	0.955	0.959	0.965	0.974
13 prightsb	0.977	0.950	0.943	0.942	0.923
32 EngFrac	0.860	0.933	0.919	0.908	0.874
14 civlibb	0.898	0.917	0.922	0.907	0.897
23 CATH	0.884	0.855	0.855	0.841	0.807
11 revcoup	0.871	0.905	0.905	0.874	0.834
31 OthFrac	0.979	0.882	0.875	0.831	0.766
15 ABSLATIT	0.914	0.874	0.821	0.851	0.743
16 AVELF	0.851	0.843	0.853	0.819	0.817
30 EcOrg	0.787	0.855	0.849	0.856	0.859
8 dpop6090	0.735	0.796	0.795	0.797	0.832
25 HINDU	0.869	0.716	0.729	0.716	0.681
6 OECD	0.828	0.647	0.579	0.518	0.618
21 SPAIN	0.878	0.644	0.586	0.701	0.794
19 BRIT	0.600	0.544	0.566	0.516	0.579
26 JEW	0.813	0.677	0.670	0.641	0.586
20 FRENCH	0.641	0.502	0.512	0.537	0.651
9 h60	0.527	0.503	0.544	0.524	0.580
Posterior Mean Model Size	9.00	9.89	10.68	11.73	14.19

Note: The left hand side variable in all regressions is the growth rate from 1960-1992 across 98 countries. Each column contains the sign-certainty probability of the regression slope coefficient for the given variable. This measures our posterior degree of confidence that the signs of the conditional means in Table 4 are correct. These are calculated with the same data but which different prior mean model sizes as labeled in the column headings. They are based on different random samples of all possible regressions using the same convergence criterion for stopping sampling. Samples range from around 4 million regressions for $\bar{k}=5$ to around 30 million for $\bar{k}=16$.

Figure 1: Posterior Distributions

These figures show the posterior distributions of the marginal effects of each independent variable on long-run economic growth. The rectangle centered on zero represents the probability mass at zero, the posterior probability that the variable is *not* contained in the model.
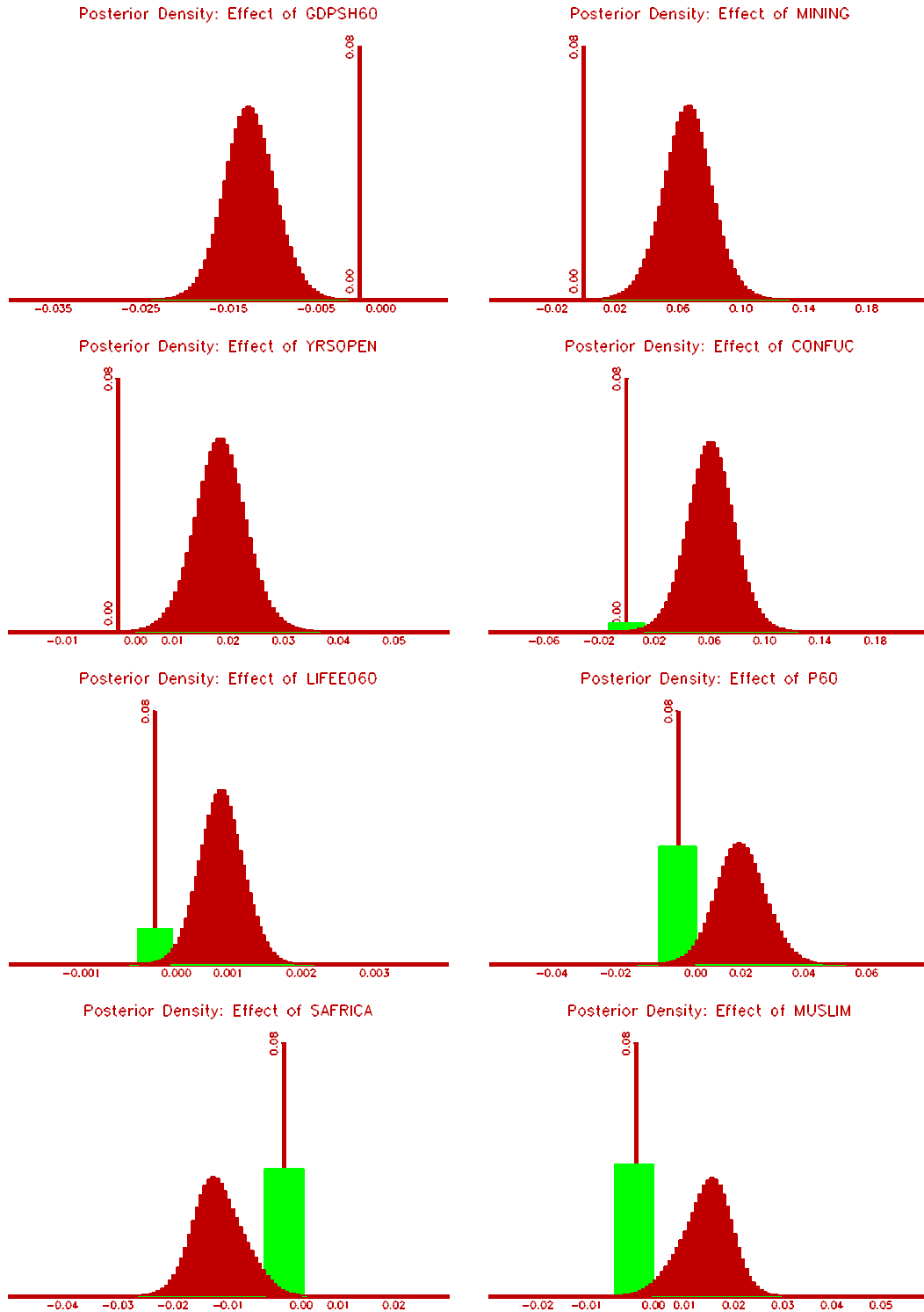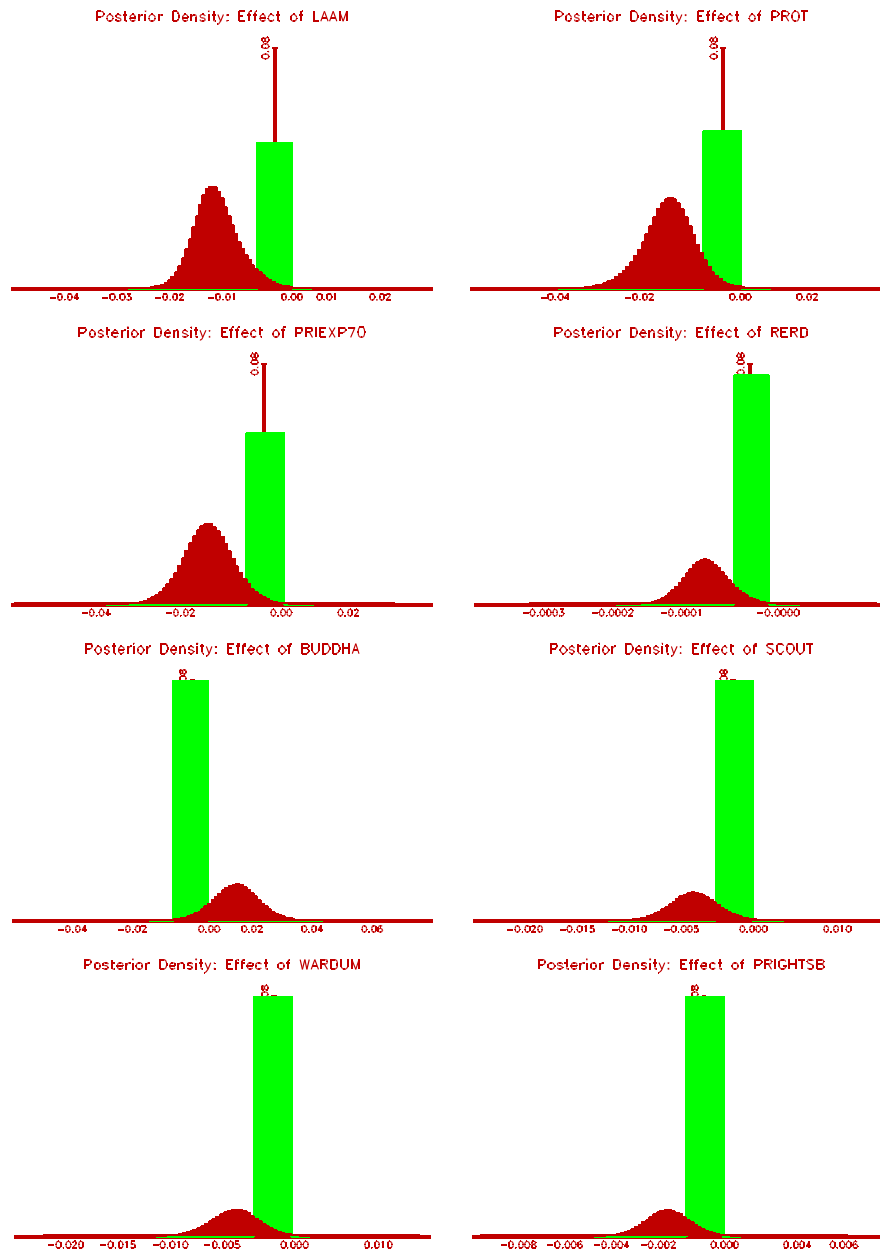

49

Figure 1 (continued): Posterior Distributions

Figure 1 (continued): Posterior Distributions

Posterior Density: Effect of ENGFRAC

Posterior Density: Effect of CIVLIBB

Posterior Density: Effect of CATH

Posterior Density: Effect of REVCOUP

Posterior Density: Effect of OTHFRAC

Posterior Density: Effect of ABSLATIT
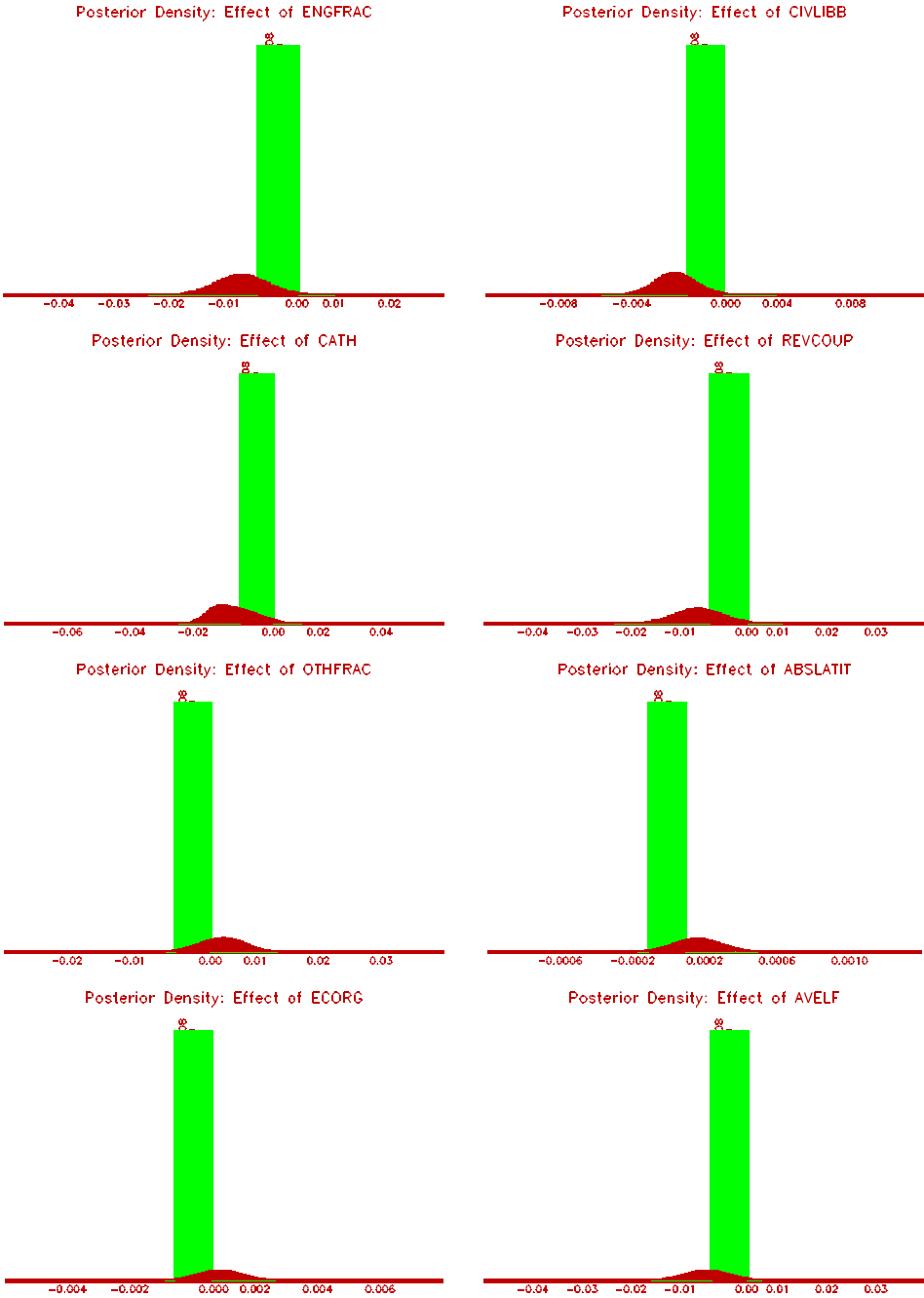
Posterior Density: Effect of ECORG

Posterior Density: Effect of AVELF

Figure 1 (continued): Posterior Distributions