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

Xavier Sala-i-Martin, Gernot Doppelhofer and Ronald I. Miller*

February 15, 2003

Abstract: This paper examines the robustness and joint interaction 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 model selection criterion. Of 67 explanatory variables we find 18 to be robustly partially correlated with long-term growth and another three variables to be marginally related. Of all the variables considered, the strongest evidence is for the relative price of investment, primary school enrolment and the initial level of real GDP per capita.

Keywords: Economic Growth, Growth Regressions, Jointness of Empirical Determinants of Economic Growth, Model Averaging

JEL Classification: O51, O52, O53.

* Sala-I-Martin: Department of Economics, Columbia University, New York NY-10027, and Universitat Pompeu Fabra (e-mail: xs23@columbia.edu); Doppelhofer: Faculty of Economics and Politics, University of Cambridge, Cambridge CB3 9DD, UK, and Trinity College, Cambridge (e-mail: gd237@cam.ac.uk); Miller: National Economic Research Associates, 1166 Avenue of the Americas, New York, NY 10036 (e-mail: ronald.miller@nera.com). We thank Manuel Arellano, Robert J. Barro, Steven Durlauf, Sanket Mohapatra, Chris Sims, Arnold Zellner, an anonymous referee and participants to the CREI Euroconference on Innovation and Growth at Universitat Pompeu Fabra for their comments. We also thank ISETR at Columbia University for allowing us to use their computer facilities. Xavier Sala-i-Martin acknowledges the NSF grant #20321600079447 and Spain's Ministerio de Ciencia y Tecnología throught grant SEC2001-0674 for financial support.

Following the seminal work of Robert J. 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]

$$\gamma = \alpha + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots + \beta_n \cdot x_n + \epsilon \tag{1}$$

where γ is the vector of rates of economic growth, α is a constant, and $x_1, \ldots, x_n$ 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 parameter "*A*" in the typical production function, $Y=F(K,L,A)$] is an important determinant of

[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, Nazrul Islam (1995), Francesco Caselli, Gerardo Esquivel, and Fernando Leffort (1996) or Barro and Xavier Sala-i-Martin (1995).

growth, at least along a 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. Moreover, many of these potential theories are not mutually exclusive (for example, education could very well be an important determinant of long term growth, and this does not imply that financial development is unimportant).

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

[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.

of having a sample size that allows us to include all potential regressors. Cross-country regressions provide perhaps the most extreme example: the number of proposed regressors 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 Ross E. Levine and David Renelt (1992). They applied a modified[3] version of Edward E. 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:

$$\gamma = \alpha_j + \beta_{yj} \cdot y + \beta_{zj} \cdot z + \beta_{xj} \cdot x_j + \epsilon \qquad \textbf{(2)}$$

[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.

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 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

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

and some solutions have been proposed such as the *reasonable extreme bounds* of Clive W.J. Granger and Harald F. Uhlig (1990).[5]

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, he 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 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 Harold Jeffreys (1961), although fleshed out by Leamer (1978). In this paper,

---

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

[6] The fixed regressors in Sala-i-Martin are *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.

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 Adrian E. Raftery, David Madigan and Jennifer A. Hoeting (1997) and Jeremy C. York, Madigan, Ivar Heuch and Rolf T. Lie (1995).[8] In the growth context, Carmen Fernandez, Fernando Rey and Mark J.F. Steel (2001) 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.

---

[7] Although the computational burden of our procedure is not insignificant it can be executed on a current PC. The baseline GAUSS routine for the results in table 2 takes 0.0015 seconds per regression.

[8] A summary of much of the recent work can be found in Hoeting, Madigan, Raftery and Chris T. 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.

In this paper we will use the Bayesian approach to averaging across models, while following the classical spirit of most of the empirical growth literature. We propose a model-averaging 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 size $\bar{k}$. 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.

[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".

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.”

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 eighteen variables which are strongly related to economic growth. They have a great deal of explanatory power and are very precisely estimated. Another three variables are marginal: they would be reasonable regressors if a researcher had a strong prior belief in their relevance. The remaining forty six variables have weak explanatory power and are imprecisely estimated.

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

I.- STATISTICAL THEORY

A. 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 Dale J. Poirier (1995). We begin with Bayes’ rule. This is simply a theorem, a basic consequence of conditional probability. Bayes’ rule in densities is:

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

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:

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

where $P(M_i|y)$ 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.

B. Diffuse Priors

As we explained above, 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

(5)

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

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 (see, for example, Arnold 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 precision matrices as $A = gX'X$ and $B = gZ'Z$ (the g-priors suggested by Zellner, 1986) and take the limit of (5) as g goes to zero we get:

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

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, $X'X$ and Z'Z, become 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

[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.

information. If we assume that the variance-covariance matrix of the X's exists and take the limit of (5) as $X'X$ (and Z'Z, respectively) goes to infinity we get:[12]

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

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 model selection criteria for the two models, exponentiated. The similarity to the Schwarz criterion is not coincidental: Gideon 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:

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

---

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

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

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 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:

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

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:

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

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

---

[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.

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* 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.

C. 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

[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 think it less likely that another measure, such as the number of assassinations, be included as well. While this sort of interdependence can be readily incorporated into our framework, we do not presently pursue this avenue.

that most researchers prefer relatively modest parameterizations. We examine the robustness of our conclusions with respect to this hyper-parameter in Section III.B.

In order to illustrate further this issue, the figure 1 plots the prior probability distributions by model size for our baseline model with $\bar{k}$=7 and with equal probabilities for all models, $\bar{k}$=33, given the 67 potential regressors we consider in our empirical work. Note that in the latter case, the great majority of the prior probability is focused on models with many included variables: more than 99 percent of the prior probability is located in models with twenty five or more included variables. It is our strong opinion that few researchers actually have such prior beliefs. Thus while we will calculate results for large models ($\bar{k}$=22 and 28) below, we do not choose to focus attention on this case.

D. 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 = 67, this would mean estimating 1.48×10^{20} regressions, which is computationally not feasible. As a result, only a relatively small subset of the total number of possible regressions can be run.

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 (Edward I. George and Robert E. McCulloch, 1993) and the Gibb's sampler-based method of John F. 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[16].

In contrast we will take a simpler approach that matches the form of the prior distribution. We select models 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.[17] Merlise Clyde, Heather Desimone, and Giovanni 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 convergence. Appendix 1 discusses computational and convergence issues in detail and may be of interest to researchers looking to apply these techniques.

---

[16] For a recent survey of computationally intensive methods see Geweke and Michael Keane (2001).

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

As an alternative to model averaging, Leamer (1978) suggests orthogonalizing[18] the explanatory variables and estimating the posterior means of the effects of the K+1 principle components. An advantage of this approach is the large reduction in the computational burden, which is especially relevant for prediction. The problem, however, is that with the transformation of the data the economic interpretation of the coefficients associated with the original variables is lost. As we are particularly interested in the interpretation of variables as determinants of economic growth, we do not follow this approach.

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 67 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 Jong-Wha Lee (1995), we only use the values for 1960. We also exclude 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 Stephen Knack and Philip Keefer's bureaucracy and corruption variables, which were computed for 1985

[18] An orthogonalization of the explanatory variables would make BACE results invariant with respect to linear transformations of the data (see also the discussion in Leamer, 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 miss observations for different countries, we selected the 67 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 68 variables (including the dependent variable, the growth rate of GDP per capita between 1960 and 1996) for 88 countries. The variable names, 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 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 I.B, which provides a summary measure of how much the data favor the inclusion of a particular variable in the growth regressions. Figure 2 shows the posterior densities (approximated by histograms) of the coefficient estimates for four selected variables (the *investment price*, the *initial level of GDP per capita*, *Primary Schooling*, and the *number of years an economy has been "open"*).[19] Note that, in Figure 2, 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[20]. As described in section I, these densities are weighted averages of the posterior densities conditional on each particular model with the weights given by the posterior model probabilities. A standard result from Bayesian theory (see, e.g., Leamer, 1978; or Poirier, 1995) is that if priors are taken as diffuse by taking the limit of a Normal-Student prior[21] then the posterior can be represented by:

---

[19] The figures for the remaining variables are available from the authors upon request.

[20] 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 0.05 meaning that for Years Open with relatively low inclusion probability of 0.09 the rectangle shows less probability mass than it actually has. All of the figures are plotted with the same vertical axis scaling.

[21] 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.

$$t_i = \frac{\beta_i - \hat{\beta}_i}{s[(X'X)^{-1}]_{ii}} \sim t(T-k) \tag{11}$$

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 2 look reasonably Gaussian.

A. Baseline Estimation

This section presents the baseline estimation results with a prior expected model size, $\bar{k} = 7$. The choice of the baseline model size is motivated by the fact that most empirical growth studies include a moderate number of explanatory variables. The posterior model size for the baseline estimation is 7.46, which is very close to the prior model size. In Section III.B we check the robustness of our results to changes in the prior mean model size. The results are based on approximately 89 million randomly drawn regressions[22].

[22] The total number of possible regression models equals 2^{67}, which is approximately equal to 1.48×10^{20} models. However, convergence of the estimates is attained relatively quickly; after 33 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^{-3} per 10,000 , and after 89 million draws the maximum change is smaller than 10^{-6}. The latter tolerance was used as one of the convergence criteria for the reported estimates. See Appendix 1 for further details.

Table 2 shows the results for the top 38 variables:[23] Column (1) reports the posterior inclusion probability of a variable in the growth regression. Variables are sorted in descending order of this posterior probability. 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 model selection 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.

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. Since our expected model size is 7, the prior inclusion probability is 7/67 = 0.104. There are 18 variables for which the posterior inclusion probability increases (these variables are shaded in Table 2). For these variables, our belief that they belong in the regression is strengthened once we see the data. We could label these variables as "strong" or "robust". The remaining 49 variables have little or no support for inclusion: seeing the data further reduces our already modest initial assessment of their inclusion probability.

Columns (2) and (3) show the posterior mean and standard deviation of the distributions, conditional on the variable being included in the model. That is, these are the means and

[23] We only report the results for the top variables because of space constraints. The results for all other variables are available from the authors on request.

standard deviation of the "hump-shaped" part of the distribution shown in figure 2. The true (unconditional) posterior mean is computed according to equation (9) while the posterior standard deviation is the square root of the variance formula in equation (10). The true 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. Hence, the unconditional posterior mean can be easily computed by the reader by multiplying the conditional mean in column (2) times the posterior inclusion probability in column 1.[24]

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.[25] 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 7/67 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

---

[24] Similarly, the unconditional variance can be calculated from the conditional variance as follows:

(13) $$\sigma_{uncond.}^{2} = [\sigma_{cond.}^{2} + \beta_{cond.}^{2}] * PosteriorInclusionProb. - \beta_{uncond.}^{2}$$

[25] 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.

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.[26]

From the posterior density we can also estimate the posterior probability, conditional on inclusion, that a coefficient has the same sign as its posterior mean[27]. This "sign certainty probability", reported in column (4), 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 percent significant in a two-sided test if 97.5 percent 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 percent significant in every single regression its sign certainty probability would be 97.5 percent. Applying a 0.975 cutoff to this quantity identifies a set of 13 variables, all of which are also in the group of 18 "strong" variables for which the posterior inclusion probability is larger than the prior inclusion probability. The remaining five have very large sign certainty probabilities (between 0.970 and 0.975.) Note that there is in principle no reason why a variable could not have a very high

---

[26] 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 percent Bayesian coverage region that excludes zero.

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

posterior inclusion probability and still have a low sign certainty probability. It just happens that in our dataset there are no such variables.[28]

The final column (5) in table 2 shows the fraction of regressions in which the variable is classically significant at the 95 percent level, in the sense of having a *t*-statistic with an absolute value greater than two. This is separated from the rest of the table because it was calculated separately from the other estimates.[29] This is reported partly for sake of comparison with extreme bounds analysis results. Note that for all the variables, many individual regressions can be found in which the variable is not significant, but even the top 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 and the posterior mean model size is 7.46. This number is, of course, sensitive to the specification of the prior mean model size, as we discuss below.

We are now ready to analyze the variables that are "strongly" related to growth.

Variables Strongly or Robustly Related to Growth

Not surprisingly, the top variable is the *Dummy for East Asian countries,* which is positively related with economic growth. This, of course, reflects the exceptional growth

---

[28] 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. Notice that the BACE weights in equation (8) penalize the inclusion of additional variables that are strongly correlated with other included regressors and do not explain more of the variation of the dependent variable.

[29] This column was calculated based on a run of 72.5 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.

performance of East Asian countries between 1960 and the mid 1990s. Notice that the dummy is present despite the robust positive relationship between the *Fraction of population Confucian* (which is ranked 9th in the table). Although the Confucius variable can be interpreted as a dummy for East Asian economies, it gains relative to the East Asia dummy variable as more regressors are included in the regression with larger prior model sizes (this is seen in table 3). The posterior mean coefficient is very precisely estimated to be positive. The sign certainty probability in column (4) shows that the probability mass of the density to the left of zero equals to 0.9992. Notice that the fraction of regressions for which the East Asian dummy has a *t*-statistic greater than two in absolute value is 99 percent.

The second variable is a measure of human capital: the *primary schooling enrolment rate in 1960* . This variable is positively related to growth and the inclusion probability is 0.80. The posterior distribution of the coefficient estimates is shown in the first panel of figure 2. Since the inclusion probability is relatively high, the mass at zero (which shows one minus this inclusion probability) is relatively small. Conditional on being included in the model, a ten percentage point increase of the primary school enrolment rate is associated with a 0.27 percentage point increase of the growth rate. This can be contrasted with the average sample growth rate of 1.82 percent between 1960 and 1996. The sign certainty probability for this variable is also 0.999 and the fraction of regressions with a *t*-statistic larger than two is 96 percent.

The third variable is the *average price of investment goods between 1960 and 1964*. Its inclusion probability is 0.77. This variable is also depicted graphically in the second panel of Figure 2. The posterior mean coefficient is very precisely estimated to be negative, which indicates that a relative high price of investment goods at the beginning of the sample is strongly and negatively related to subsequent income growth[30]. The sign certainty probability in column (4) shows that the probability mass of the density to the left of zero equals to 0.99: this can also be seen in Figure 2 by the fact that almost all of the continuous density lies below zero.

The next variable is the *initial level of per capita GDP* , a measure of conditional convergence. The inclusion probability is 0.69. The third panel in Figure 2 shows the posterior distribution of the coefficient estimates for initial income. Conditional on inclusion, the posterior mean coefficient is -0.009 (with a standard deviation of 0.003). In other words, the coefficient associated with conditional convergence is very precisely estimated, although the mean coefficient is somewhat smaller than the convergence coefficient reported in the economic growth literature. The sign certainty probability in column (4) shows that the probability mass of the density to the left of zero equals 0.999. The fraction of regressions in which the coefficient for initial income has *t*-statistic greater than two in absolute value is only 30 percent, so that an extreme bounds test very easily labels the variable as not robust. Nonetheless, the tightly 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, so they receive little weight in the averaging process. Furthermore, the high inclusion probability suggests that regressions that omit initial income are likely to perform poorly.

---

[30] Once the relative price of investment goods is included among the pool of explanatory variables, the share of investment to GDP in 1961 becomes insignificant and has the "wrong sign" while the other results are unaffected. The estimation results including investment share are available from the authors upon request.

The next variables reflect the poor economic performance of tropical countries: The *Proportion of a Country's Area in the Tropics* has a negative relationship with income growth. Similarly, the *Index of Malaria Prevalence* has a negative relationship with income growth indicating that unfavourable geographic conditions or exposure to the Malaria virus are associated with lower growth. Table 3 shows that the posterior inclusion probability of the Malaria index falls as models become larger indicating that it could act as catch-all variable when few explanatory variables are included, but drops in significance as more variables explaining different steady states enter.

Another geographical variable that performs well is the *Density of the Population in Coastal Areas*, which has a positive relationship with growth suggesting that areas which are densely populated and are close to the sea have experienced higher growth rates.

The negative growth performance of malaria ridden countries is an indication that the health component of human capital is an important determinant of growth. Another variable that reflects this phenomenon is the *Life expectancy in 1960*, which reflects things like nutrition, health care, or literacy rates. Countries with high life expectancy in 1960 tended to grow faster over the following four decades. The inclusion probability for this variable is 0.28.

Dummies for *Sub-Saharan Africa* and *Latin America* are negatively related to income growth. The posterior means conditional on inclusion are both negative, implying that Latin American and Sub-Saharan African countries had income per capita growth rates between 1960 and 1996 which were 1.47 and 1.28 percentage points respectively below the level that would be predicted by the countries' other characteristics. For comparison, the sample average growth rate is 1.82. The African dummy is significant in 90 percent of the regressions and the sign certainty probability is 98 percent. Although the Latin American dummy is only significant in 33 percent of the regressions, its sign certainty is almost as high as the African: 97 percent.

The *Fraction of GDP in Mining* has a positive relationship with growth and inclusion probability of 0.12. This variable captures the success of countries with a large endowment of natural resources. Many economists expect that the large rents associated with more political instability or rent-seeking and low growth, our study shows that economies with a larger mining sector tend to perform better.[31]

Former Spanish colonies tend to grow less[32] whereas the *number of years an economy has been open* has a positive sign. Both of these variables have inclusion probabilities that increase only moderately with larger models in Table 3 indicating that they capture steady state variations in smaller models, but are relatively less important in explaining growth when more variables are included in the regression models.

Both the *fraction of the population Muslim and Buddhist* have a positive association with growth where the conditional mean of the latter variable is almost twice as large (0.022) than for the fraction Muslim (0.013), but both are relatively small compared to the fraction Confucian (0.054). The index of *ethnolinguistic fractionalization* is negatively related to growth and it also appears to be robust.

---

[31] The regressions that include the fraction of mining tend to have an outlier: Botswana, which is country that discovered diamonds in its territory in the 1960s and has managed to exploit them successfully (which implies a large share of mining in GDP) and has experienced extraordinary growth rates over the last forty years.

[32] We will abstain from analyzing the meaning of this empirical finding.

Finally, the *share of government consumption in GDP* is also robustly estimated and its sign is negative. This could be expected because public consumption does not tend to contribute to growth directly, but it needs to be financed with distortionary taxes which hurt the growth rate. Perhaps the real surprise is the negative coefficient of the *public investment share.* Table 2 shows that this variable is not robust when the prior model size is $\bar{k} = 7$. However, we will see later that this is one of the variables that becomes important in larger models and the sign remains *negative*.

Variables Marginally Related to Growth

There are three variables that have posterior probabilities somewhat lower than their prior probabilities but nonetheless are fairly precisely estimated if they are included in the growth regression (that is, their sign certainty probability is larger than 95 percent). These variables are: the overall *density* in 1960 (which is positively related to growth), *real exchange rate distortions* (negative) and the *fraction of population speaking a foreign language* (positive).

Variables Not Robustly Related to Growth

The remaining forty-six 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. It is interesting to 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 or a *Socialism dummy* measuring whether a country was significant time under Socialism, have no strong relationship with growth between 1960 and 1996.[33] This could be due to the fact that other variables, capturing political or economic instability such as the relative price of investment goods, 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.

We also notice that some macroeconomic variables such as the *inflation rate* do not appear to be strongly related to growth. Other surprisingly weak variables are the *spending in Public Education*, some measures of *higher education*, some geographical measures such as the *latitude or* the *distance from the equator*, and various proxies for "scale effects" such as the total *population*, *aggregate GDP*, or the total area of a country.

B. 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. Table 3 does precisely this, reporting the posterior inclusion probabilities and conditional posterior means, respectively, for $\bar{k}$ equal to 5, 9, 11, 16 and 22[34] as well as repeating the benchmark numbers for easy comparison. Note that each $\bar{k}$ has a corresponding value of the prior probability of inclusion, which is reported in the first row of the table. Thus, to see whether a variable improves its probability of inclusion relative to the prior, we need to compare the posterior probability to the

[33] We should remind the reader that most of the economies Eastern Europe and the former Soviet block are not included in our data set (see Appendix 2 for the list of included countries).

[34] In the not-for-publication appendix, we also report the results for larger prior model sizes.

corresponding prior probability. The variables that are important in the baseline case of $\bar{k} = 7$ and are not important for other prior model sizes are shaded in Table 3. Variables that are not important for $\bar{k} = 7$ but become important with other sizes are both shaded and their cells are bordered.

"Strong" variables that become "weak"

Note that most of the strongest variables show very little sensitivity to the choice of prior model size, either in terms of their inclusion probabilities or their coefficient estimates.[35] Some of the important variables seem to improve substantially with the prior model size. For example, for *the fraction of GDP in Mining,* the posterior inclusion probability rises from 7 percent with $\bar{k} = 5$ to 66 percent with $\bar{k} = 22$. This suggests that Mining is a variable which requires other conditioning variables in order to display its full importance[36]. Both the *fraction of Confucians* and the *Sub-Saharan Africa dummy* are also variables which appear to do better with more conditioning variables and have stable coefficient estimates.

Although most of the strong variables remain strong, five of them tend to lose power as we increase the prior model size. That is, for larger models, the posterior probability declines to levels below the prior size. These variables are the *index of malaria prevalence*, the *former Spanish colony*, the number of *years an economy has been open,* the index of *ethnolinguistic*

[35] The coefficient estimates are not reported here, but are available from the authors.

[36] Notice that the fraction of GDP in Mining is largely driven by the success of Botswana. Once other control variables such as a Sub-Saharan dummy are included, the Mining variable captures the unusual performance of such countries.

fractionalization and the *government consumption share*. This suggests that these variables could be acting as catch-all for various other effects. For example, 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). The other 13 variables that were robust in the baseline model also appear to be robust to different prior specifications.

"Weak" Variables that become "Strong"

At the other end of the scale, most of the forty-six variables that showed little partial correlation in the baseline estimation are not helped by alternative priors[37]. Their posterior inclusion probabilities rise as $\bar{k}$ increases, which is hardly surprising as their prior inclusion probabilities are rising. But their posterior inclusion probabilities remain below the prior so we are force to think of them as "weak".

There are three variables that are weak in the baseline study but become "strong" with some prior model sizes. These are the *population density,* the *fraction of population that speak a foreign language* (a measure of international social capital and openness) and the *public investment share.* As mentioned above, the public investment share is particularly interesting because it becomes strong for larger prior model sizes, but the sign of the correlation is *negative*. That is, a larger public investment share tends to be associated with lower grow rates.

[37] The exception is the *public investment share* which has a posterior inclusion probability greater than the prior for relatively large models with $\bar{k} \geq 16$.

Our interpretation of these results is that our baseline results are very robust to alternative prior size specifications. This robustness applies also to the "sign certainty probabilities," which are not reported here because of space constraints, but are available from the authors on request.

Nonlinearities

The literature has identified some variables that 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. To test this hypothesis, we include the average inflation rate in the 1960s, 1970s and 1980s and its square as separate regressors. The BACE procedure allows such variables to enter individually and the data would assign larger weight to well-fitting models if there was a non-linear relationship. The posterior inclusion probabilities for inflation and its square are very low and the conditional coefficient estimates are not different from zero.

Jointness of Growth Determinants

In addition to looking at the inclusion probabilities of individual variables, we can also investigate the jointness of regressors in explaining economic growth. The jointness statistic reported in table 4[38] are based on the joint inclusion probabilties of 2 variables x_i and x_j. The entries of table 4 are calculated by dividing the joint inclusion probabilities by the product of the individual inclusion probability for each variable separately. If two variables were independent in the posterior (as they are in the prior), this number would be one. Numbers greater than one

[38] Note that table 4 reports only the jointness statistic for the top 21 variables. The statistic for the remaining variables is available on request; the total number of entries is very large 67x67 = 4489.

indicate variables that help each other in explaining growth (complements), values less than one indicate that variables that explain similar aspects of the dependent variable (substitutes). The index can also be interpreted as measure of collinearity between variables.

Looking at the leading variables, there is considerable variation. Some, like the *investment price index*, and the *initial income*, have very little interesting jointness. This finding is surprising for initial income since there is no absolute convergence. It seems that as long as ``some” other good variables are included, it doesn't seem to matter much which ones.

Unsurprisingly, the *East Asian* variable does not want to be included with the Fraction *Buddhist* and *Confucian* which are measuring similar aspects of the data. More surprisingly, the East Asian dummy also exhibits negative jointness with the *Latin America*, *European* or *Sub Saharan African* dummies. On the other hand the Fraction Confucian has a positive association with other regional dummies. This suggests that especially regression models with few variables capture a lot of variation of growth rates with just the East Asian dummy included. On the other hand, if the East Asian dummy is not included in the model, the data askfor a full set of the other dummies (for instance, Latin America and Sub Saharan African have a strong positive joint association when included in the model.)

The *Coastal density* and *Tropical area* variables both have a complicated set of interactions. The coastal density variable has a negative association with any of the *Openness measures*. This finding seems very interesting given that the coastal density variable is robustly related with economic growth. It is suggestive that openness measures may really be picking up countries with historic opportunities for trade. Also, this variable has a strong positive interaction with the tropical area and negative interactions with the regional dummies apart from East Asia.

The tropical area, unsurprsisingly, has a negative relationship with other variables that measuring similar things, such as *Malaria Prevalence* and the *Fraction of Population in Tropics*. It also interacts negatively with the regional variables.

5.- CONCLUSIONS

In this paper we propose a Bayesian Averaging of Classical Estimates (BACE) 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 use an averaging method that is fully justified on Bayesian grounds and we do not restrict the number of 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 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 fifth of the 67 variables used in the analysis can be said to be robustly related to growth while several more are marginally related. The strongest evidence is found for primary schooling enrolment, the relative price of investment goods,, and the initial level of income where the latter reflects the concept of conditional convergence. Other important variables include regional dummies (such as East Asia, Sub-Saharan Africa or Latin America), some measures of human capital and health (such as life expectancy, proportion of a country lying in the Tropics, and Malaria prevalence), religious dummies (such as the fraction Confucians, Buddhist and Muslims) and some sectoral variables such as Mining. The public consumption and public investment shares are negatively related to growth, although the results are strong only for certain prior model sizes.

Finally, we show that our results are quite robust to the choice of the prior model size: most of the “strong” variables in the baseline case, remain strong for other prior model sizes. Our method also allows us to investigate interesting jointness in determinants of economic growth which sheds light on some interesting underlying relationships between the explanatory variables.

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:

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

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:

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

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 $t(T\text{-}k\text{-}1)$ 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:

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

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. In addition, we also checked for convergence of the posterior inclusion probability with changes less than 10E-4. 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 $\bar{k} = 7$ 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 around 80-90 million regressions; in addition we ran a sampler with 200 million stratified draws. 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

89 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 67. 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

The following 88 countries are included in the regressions:

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

REFERENCES

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

_______. "Determinants of Democracy", *Journal of Political Economy*, 1999, 107 (6), pp. S158-83.

Barro, Robert J. and Lee, Jong-Wha. "International Comparisons of Educational Attainment," *Journal of Monetary Economics*, 1993, 32 (3), pp. 363-394. The data for this paper were taken from the NBER Web Page, http://www.nber.org/pub/barro.lee/.

Barro, Robert J and Sala-i-Martin, **Xavier.** *Economic Growth.* New York: McGraw Hill, 1995.

Caselli, Francesco; Esquivel, Gerardo and Lefort, Fernando. "Reopening the Convergence Debate: a New Look at Cross-Country Growth Empirics," *Journal of Economic Growth,* 1996, 1(3), pp. 363-89.

Clyde, Merlise; Desimone, Heather and Parmigiani, Giovanni . "Prediction via Orthogonalized Model Mixing," *Journal of the American Statistical Association*, 1996, 91, pp. 1197-1208.

Doppelhofer, Gernot. "Three Essays on the Determinants of Economic Growth", *Ph.D. dissertation*, Columbia University, 2000.

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

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

Fernandez, Carmen; Ley, Eduardo and Steel, Mark F. J. "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 2001, 16(5), 563-576.

Gallup, John L.; Mellinger, Andrew and Sachs, Jeffrey D. "Geography Datasets" Center for International Development at Harvard University (CID), 2001. Data website: http://www2.cid.harvard.edu/ciddata/geographydata.htm

George, Edward I. and McCulloch, Robert E."Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 1993, 88, pp. 881-889.

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

Geweke, John and Michael Keane. "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics Vol. 5*, J.J. Heckman and E. Leamer, eds., Amsterdam, North Holland, 2001.

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

Heston, Alan; Summers, Robert and Aten, Bettina. *Penn World Table Version 6.0*. Center for International Comparisons at the University of Pennsylvania (CICUP), 2001, December. Data website: http://pwt.econ.upenn.edu/.

Hoeting, Jennifer A.; Madigan, David; Raftery, Adrian E. and Volinsky, Chris T. "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 1999, 14(4): pp. 382-417.

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

Jeffreys, Harold. *Theory of Probability, 3rd ed.* London: Oxford University Press, 1961.

Hall, Robert E. and Jones, I. Charles, "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 1996, 114(1), 83-116. Data website: http://emlab.berkeley.edu/users/chad/datasets.html.

Knack, Stephen and Keefer, Philip, "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Politics*, 1995, 7(3), pp. 207-227. The data from this paper were provided to us by Robert Barro.

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

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

___."Sensitivity Analysis Would Help," *American Economic Review*, 1985, 75 (3), pp. 308-313.

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

Madigan, David, and York, Jeremy C."Bayesian Graphical Models for Discrete Data," *International Statistical Review*, 1995, 89, pp. 215-232.

Poirier, Dale J. *Intermediate Statistics and Econometrics*. Cambridge, Mass: MIT Press, 1995.

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

Sachs, Jeffrey D. and Warner, Andrew M."Economic Reform and the Process of Economic Integration", *Brookings Papers of Economic Activity*, 1995, pp.1-95.

_______, "Natural Resource Abundance and Economic Growth", CID at Harvard University, 1997. Data website: www.cid.harvard.edu/ciddata/ciddata.html

Sala-i-Martin, X. "I Just Ran 2 Million Regressions," *American Economic Review, Papers and Proceedings*, May 1997a, 87 (2), pp. 178-183.

Sala-i-Martin, X. "I Just Ran Four Million Regressions," NBER Working Paper No. 6252, November 1997b.

Schwarz, Gideon. "Estimating the Dimension of a Model," *The Annals of Statistics*, 1978, **6**, pp. 461-464.

York, Jeremy C.; Madigan, David; Heuch, I. Ivar and Lie, Rolv Terje. "Estimating a Proportion of Birth Defects by Double Sampling: A Bayesian Approach Incorporating Covariates and Model Uncertainty," *Applied Statistics*, 1995, 44, pp. 227-242.

Zellner, Arnold . *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley and Sons, 1971.

_______. "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., Amsterdam: North-Holland, 1986, pp. 233-43.

Figure 1. Prior Probabilities by Model Size

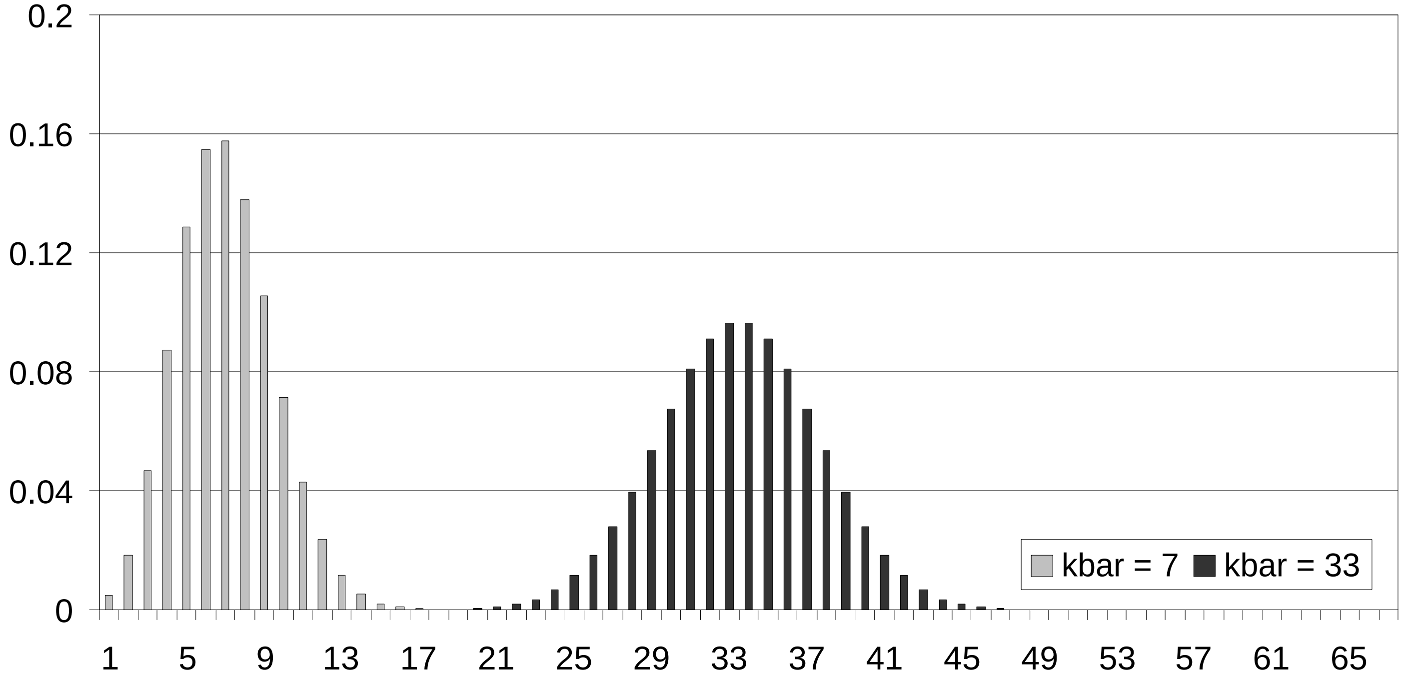
0.2
0.16
0.12
0.08
0.04
0
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
kbar = 7
kbar = 33

Figure 2. Posterior Distribution of Marginal Effects of Selected Variables on Long-Run Growth. The distribution consists of two parts: (1) the "hump-shaped" part of the distribution represents the distribution of estimates conditional on the variable being included in the model; (2) the "lump" at zero shows the posterior probability that a variable is *not* included in the regression, which equals one minus the posterior inclusion probability.

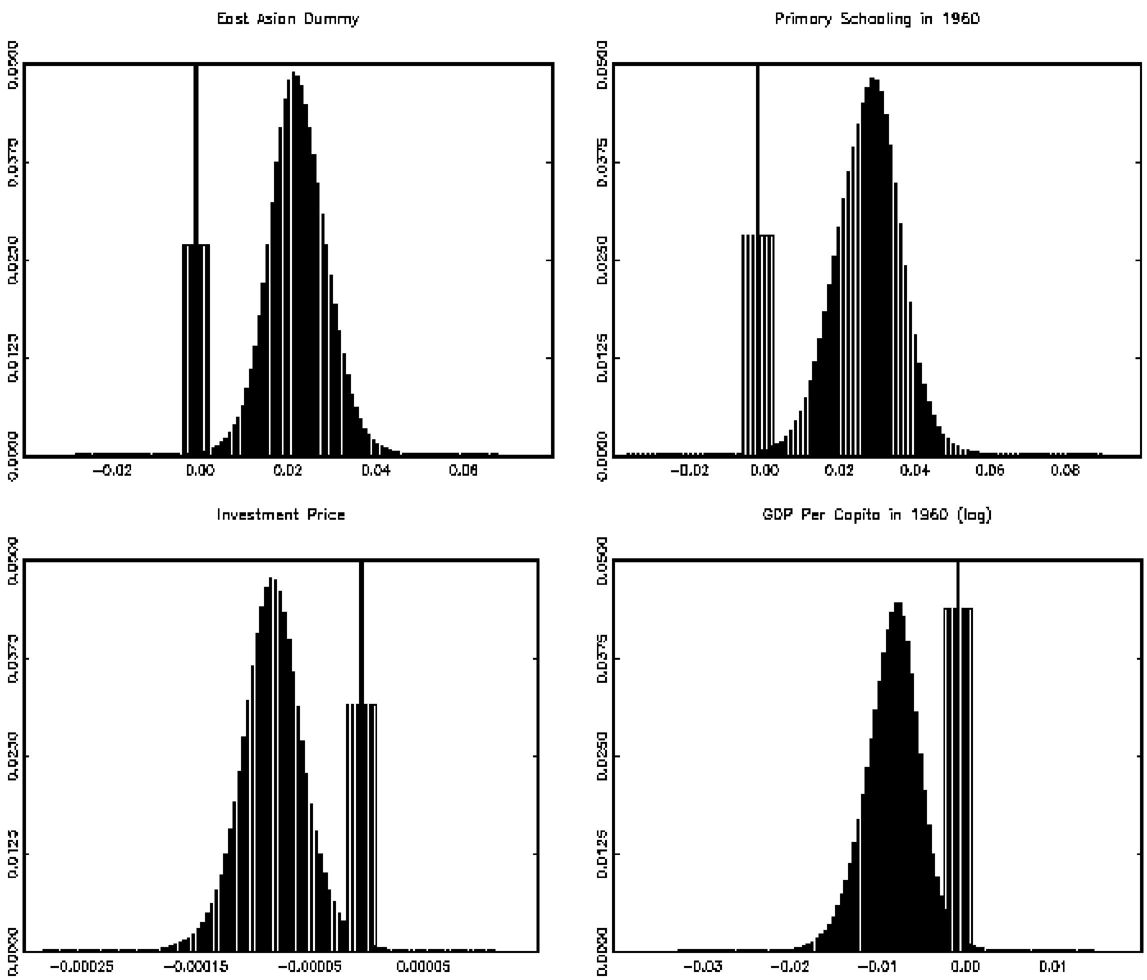

Table 1: Data Description and Sources

Rank	Variable	Description and Source	Mean	Standard Deviation
	Average Growth Rate of GDP per capita 1960-96	Growth of GDP per capita at Purchasing Power Parities between 1960 and 1996. From Heston, Summers and Aten (2001).	0.0182	0.019
1	**East Asian Dummy**	Dummy for East Asian countries.	0.1136	0.3192
2	**Primary Schooling in 1960**	Enrollment rate in primary education in 1960. Barro and Lee (1993).	0.7261	0.2932
3	**Investment Price**	Average investment price level between 1960 and 1964 on purchasing power parity basis. From Heston, Summers and Aten (2001).	92.4659	53.6778
4	**GDP in 1960 (log)**	Logarithm of GDP per capita in 1960. From Heston, Summers and Aten (2001).	7.3549	0.9011
5	**Fraction of Tropical Area**	Proportion of country's land area within geographical tropics. From Gallup, Mellinger and Sachs (2001).	0.5702	0.4716
6	**Population Density Coastal in 1960s**	Coastal (within 100km of coastline) population per coastal area in 1965. From Gallup, Mellinger and Sachs (2001).	146.8717	509.8276
7	**Malaria Prevalence in 1960s**	Index of malaria prevalence in 1966. From Gallup, Mellinger and Sachs (2001).	0.3394	0.4309
8	**Life Expectancy in 1960**	Life Expectancy in 1960. Barro and Lee (1993).	53.7159	12.0616
9	**Fraction Confucius**	Fraction of population Confucian. Barro (1999).	0.0156	0.0793
10	**African Dummy**	Dummy for sub-saharan African countries.	0.3068	0.4638
11	**Latin American Dummy**	Dummy for Latin American countries.	0.2273	0.4215
12	**Fraction GDP in Mining**	Fraction of GDP in Mining. From Hall and Jones (1999).	0.0507	0.0769
13	**Spanish Colony**	Dummy variable for fomer Spanish colonies. Barro (1999).	0.1705	0.3782
14	**Years Open 1950-94**	Number of years economy has been open between 1950 and 1994. From Sachs and Warner (1995).	0.3555	0.3444
15	**Fraction Muslim**	Fraction of population Muslim in 1960. Barro (1999).	0.1494	0.2962
16	**Fraction Buddhist**	Fraction of population Buddhist in 1960. Barro (1999).	0.0466	0.1676
17	**Ethnolinguistic Fractionalization**	Average of five different indices of ethnolinguistic fractionalization which is the probability of two random people in a country not speaking the same language. From Easterly and Levine (1997).	0.3476	0.3016
18	**Gov. Consumption Share 1960s**	Share of expenditures on government consumption to GDP in 1961. Barro and Lee (1993).	0.1161	0.0745
19	**Population Density 1960**	Population per area in 1960. Barro and Lee (1993).	108.0735	201.4449
20	**Real Exchange Rate Distortions**	Real Exchange Rate Distortions. Levine and Renelt (1992).	125.0341	41.7063
21	**Fraction Speaking Foreign Language**	Fraction of population speaking foreign language. Hall and Jones (1999).	0.3209	0.4136
22	**Openess measure 1965-74**	Ratio of exports plus imports to GDP, averaged over 1965 to 1974. This variable was provided by Robert Barro.	0.5231	0.3359
23	**Political Rights**	Political rights index. From Barro (1999).	3.8225	1.9966
24	**Government Share of GDP in 1960s**	Average share government spending to GDP between 1960-64. From Heston, Summers and Aten (2001).	0.1664	0.0712
25	**Higher Education 1960**	Enrollment rates in higher education. Barro and Lee (1993).	0.0376	0.0501
26	**Fraction Population In Tropics**	Proportion of country's population living in geographical tropics. From Gallup, Mellinger and Sachs (2001).	0.3	0.3731
27	**Primary Exports 1970**	Fraction of primary exports in total exports in 1970. From Sachs and Warner (1997).	0.7199	0.2827
28	**Public Investment Share**	Average share of expenditures on public investment as fraction of GDP between 1960 and 1965. Barro and Lee (1993).	0.0522	0.0388
29	**Fraction Protestants**	Fraction of population Protestant in 1960. Barro (1999).	0.1354	0.2851
30	**Fraction Hindus**	Fraction of the population Hindu in 1960. Barro (1999).	0.0279	0.1246
31	**Fraction Population Less than 15**	Fraction of population younger than 15 years in 1960.	0.3925	0.0749
32	**Air Distance to Big Cities**	Logarithm of minimal distance (in km) from New York, Roterdam or Tokio. From Gallup, Mellinger and Sachs (2001).	4324.1705	2613.7627
33	**Nominal Govertnment GDP Share 1960s**	Average share of nominal government spending to nominal GDP between 1960 and 1964. Calculated from Heston, Summers and Aten (2001).	0.149	0.0584
34	**Absolute Latitude**	Absolute Latitude. Barro (1999).	23.2106	16.8426
35	**Fraction Catholic**	Fraction of population Catholics in 1960. Barro (1999).	0.3283	0.4146
36	**Fertility in 1960s**	Fertility in 1960s. Barro and Sala-i-Martin (1995).	1.562	0.4193
37	**European Dummy**	Dummy for European economies.	0.2159	0.4138
38	**Outward Orientation**	Measure of outward orientation. Levine and Renelt (1992).	0.3977	0.4922
39	**Colony Dummy**	Dummy for Former Colony. Barro (1999).	0.75	0.4355
40	**Civil Liberties**	Index of Civil liberties index in 1972. Barro (1999).	0.5095	0.3259

Table 1 (cont'): Data Description and Sources

41	**Revolutions and Coups**	Number of revolutions and military coups. Barro and Lee (1993).	0.1849	0.2322
42	**British Colony Dummy**	Dummy for former British colony after 1776. Barro (1999).	0.3182	0.4684
43	**Hydrocarbon Deposits in 1993**	Log of hydrocarbon deposits in 1993. From Gallup, Mellinger and Sachs (2001).	0.4212	4.3512
44	**Fraction Population Over 65**	Fraction of population older than 65 years in 1960.	0.0488	0.029
45	**Defense Spending Share**	Average share public expenditures on defense as fraction of GDP between 1960 and 1965. Barro and Lee (1993).	0.0259	0.0246
46	**Population in 1960**	Population in 1960. Barro (1999).	20308.08	52538.3866
47	**Terms of Trade Growth in 1960s**	Growth of terms of trade in the 1960. Barro and Lee (1993).	-0.0021	0.0345
48	**Public Education Spending Share in GDP in 1960s**	Average share public expenditures on education as fraction of GDP between 1960 and 1965. Barro and Lee (1993).	0.0244	0.0096
49	**Landlocked Country Dummy**	Dummy for landlocked countries.	0.1705	0.3782
50	**Religious Intensity**	Religion measure. Barro (1999).	0.7803	0.1932
51	**Size of Economy**	Logarithm of aggregate GDP in 1960.	16.1505	1.8202
52	**Socialist Dummy**	Dummy for country under Socialist rule for considerable time during 1950 to 1995. From Gallup, Mellinger and Sachs (2001).	0.0682	0.2535
53	**English Speahing Population**	Fraction of population speaking English. From Hall and Jones (1999).	0.084	0.2522
54	**Average Inflation 1960-90**	Average inflation rate between 1960 and 1990. Levine and Renelt (1992).	13.1298	14.9899
55	**Oil Producing Country Dummy**	Dummy for oil producing country. Barro (1999).	0.0568	0.2328
56	**Population Growth Rate 1960-90**	Average growth rate of population between 1960 and 1990. Barro and Lee (1993).	0.0215	0.0095
57	**Timing of Independence**	Timing of national independence measure: 0 if before 1914; 1 if between 1914 and 1945; 2 if between 1946 and 1989; and 3 if after 1089. From Gallup, Mellinger and Sachs (2001).	1.0114	0.9767
58	**Fraction of Land Area Near Navigable Water**	Proportion of country's land area within 100km of ocean or ocean-navigable river. From Gallup, Mellinger and Sachs (2001).	0.4722	0.3802
59	**Square of Inflation 1960-90**	Square of average inflation rate between 1960 and 1990.	394.5368	1119.6992
60	**Fraction Spent in War 1960-90**	Fraction of time spent in war between 1960 and 1990. Source: Barro and Lee (1993).	0.0695	0.1524
61	**Land Area**	Area in km2. Barro and Lee (1993).	867188.52	1814688.29
62	**Tropical Climate Zone**	Fraction Tropical Climate Zone. From Gallup, Mellinger and Sachs (2001).	0.19	0.2687
63	**Terms of Trade Ranking**	Terms of trade ranking. Barro (1999)	0.2813	0.1904
64	**Capitalism**	Degree Capitalism Index. From Hall and Jones (1999).	3.4659	1.3809
65	**Fraction Othodox**	Fraction of population Orthodox in 1960. Barro (1999).	0.0187	0.0983
66	**War Particpation 1960-90**	Indicator for countries that participated extrenal war between 1960 and 1990. Barro and Lee (1993).	0.3977	0.4922
67	**Interior Density**	Interior (more than 100 km from coastline) population per interior area in 1965. From Gallup, Mellinger and Sachs (2001).	43.3709	88.0626

Table 2: Baseline Estimation

Rank	Variable	Posterior Inclusion Probabiliy	Posterior Mean Conditional on Inclusion	Posterior s.d. Conditional on Inclusion	Sign Certainty Probability	Fraction of Regressions with |tstat|>2
		(1)	(2)	(3)	(4)	(5)
1	**East Asian**	0.823	0.021805	0.006118	0.999	0.99
2	**Primary Schooling 1960**	0.796	0.026852	0.007977	0.999	0.96
3	**Investment Price**	0.774	-0.000084	0.000025	0.999	0.99
4	**GDP 1960 (log)**	0.685	-0.008538	0.002888	0.999	0.30
5	**Fraction of Tropical Area (or people)**	0.563	-0.014757	0.004227	0.997	0.59
6	**Pop. Density Coastal 1960s**	0.428	0.000009	0.000003	0.996	0.85
7	**Malaria Prevalence in 1960s**	0.252	-0.015702	0.006177	0.990	0.84
8	**Life Expectancy in 1960**	0.209	0.000808	0.000354	0.986	0.79
9	**Fraction Confucious**	0.206	0.054429	0.022426	0.988	0.97
10	**African Dummy**	0.154	-0.014706	0.006866	0.980	0.90
11	**Latin American Dummy**	0.149	-0.012758	0.005834	0.969	0.30
12	**Fraction GDP in Mining**	0.124	0.038823	0.019255	0.978	0.07
13	**Spanish Colony**	0.123	-0.010720	0.005041	0.972	0.24
14	**Years Open**	0.119	0.012209	0.006287	0.977	0.98
15	**Fraction Muslim**	0.114	0.012629	0.006257	0.973	0.11
16	**Fraction Buddhist**	0.108	0.021667	0.010722	0.974	0.90
17	**Ethnolinguistic Fractionaliz.**	0.105	-0.011281	0.005835	0.974	0.52
18	**Gov. Consumption Share 60s**	0.104	-0.044171	0.025383	0.975	0.77
19	**Population Density 1960s**	0.086	0.000013	0.000007	0.965	0.01
20	**Real Exc. Rate Distortions**	0.082	-0.000079	0.000043	0.966	0.92
21	**Fraction Speaking Foreign Language**	0.080	0.007006	0.003960	0.962	0.43
22	**(Imports + Exports)/GDP**	0.076	0.008858	0.005210	0.949	0.67
23	**Political Rights**	0.066	-0.001847	0.001202	0.939	0.35
24	**Government Share of GDP**	0.063	-0.034874	0.029379	0.935	0.58
25	**Higher Education in 1960**	0.061	-0.069693	0.041833	0.946	0.10
26	**Fraction Popul. In Tropics**	0.058	-0.010741	0.006754	0.940	0.85
27	**Primary Exports in 1970**	0.053	-0.011343	0.007520	0.926	0.75
28	**Public Investment Share**	0.048	-0.061540	0.042950	0.922	0.00
29	**Fraction Protestants**	0.046	-0.011872	0.009288	0.909	0.29
30	**Fraction Hindus**	0.045	0.017558	0.012575	0.915	0.07
31	**Fraction Popul. Less than 15**	0.041	0.044962	0.041100	0.871	0.24
32	**Air Distance to Big Cities**	0.039	-0.000001	0.000001	0.888	0.18
33	**Gov C Share deflated with GDP prices**	0.036	-0.033647	0.027365	0.893	0.05
34	**Absolute Lattitude**	0.033	0.000136	0.000233	0.737	0.37
35	**Fraction Catholic**	0.033	-0.008415	0.008478	0.837	0.16
36	**Fertility Rates in 1960s**	0.031	-0.007525	0.010113	0.767	0.46
37	**European Dummy**	0.030	-0.002278	0.010487	0.544	0.19
38	**Outward Orientation**	0.030	-0.003296	0.002727	0.886	0.01

Table 3: Posterior Inclusion Probabilities with Different Prior Model Sizes

Rank		kbar5	kbar7	kbar9	kbar11	kbar16	kbar22
	Prior Inclusion Probability	0.075	0.104	0.134	0.164	0.239	0.328
1	**East Asian**	0.891	0.823	0.757	0.711	0.585	0.481
2	**Primary Schooling 1960**	0.709	0.796	0.826	0.862	0.890	0.924
3	**Investment Price**	0.635	0.774	0.840	0.891	0.936	0.968
4	**GDP 1960 (log)**	0.526	0.685	0.788	0.843	0.920	0.960
5	**Fraction of Tropical Area**	0.536	0.563	0.548	0.542	0.462	0.399
6	**Pop. Density Coastal 1960s**	0.350	0.428	0.463	0.473	0.433	0.389
7	**Malaria Prevalence in 1960s**	0.339	0.252	0.203	0.176	0.145	0.131
8	**Life Expectancy in 1960**	0.176	0.209	0.262	0.278	0.368	0.440
9	**Fraction Confucious**	0.140	0.206	0.272	0.333	0.501	0.671
10	**African Dummy**	0.095	0.154	0.223	0.272	0.406	0.519
11	**Latin American Dummy**	0.101	0.149	0.205	0.240	0.340	0.413
12	**Fraction GDP in Mining**	0.072	0.124	0.209	0.275	0.478	0.659
13	**Spanish Colony**	0.130	0.123	0.119	0.116	0.124	0.148
14	**Years Open**	0.090	0.119	0.124	0.132	0.145	0.155
15	**Fraction Muslim**	0.078	0.114	0.150	0.178	0.267	0.366
16	**Fraction Buddhist**	0.073	0.108	0.152	0.190	0.320	0.465
17	**Ethnolinguistic Fractionaliz.**	0.080	0.105	0.131	0.140	0.155	0.160
18	**Gov. Consumption Share 60s**	0.090	0.104	0.135	0.147	0.213	0.262
19	**Population Density 1960s**	0.043	0.086	0.137	0.175	0.257	0.295
20	**Real Exc. Rate Distortions**	0.059	0.082	0.117	0.134	0.205	0.263
21	**Fraction Speaking Foreign Language**	0.052	0.080	0.110	0.149	0.247	0.374
22	**(Imports + Exports)/GDP**	0.063	0.076	0.085	0.099	0.131	0.181
23	**Political Rights**	0.042	0.066	0.082	0.095	0.114	0.130
24	**Government Share of GDP**	0.044	0.063	0.087	0.112	0.186	0.252
25	**Higher Education in 1960**	0.059	0.061	0.066	0.070	0.079	0.103
26	**Fraction Popul. In Tropics**	0.047	0.058	0.061	0.074	0.099	0.132
27	**Primary Exports in 1970**	0.047	0.053	0.065	0.072	0.104	0.137
28	**Public Investment Share**	0.023	0.048	0.096	0.151	0.321	0.525
29	**Fraction Protestants**	0.035	0.046	0.055	0.061	0.083	0.120
30	**Fraction Hindus**	0.028	0.045	0.059	0.077	0.126	0.179
31	**Fraction Popul. Less than 15**	0.035	0.041	0.045	0.050	0.067	0.093
32	**Air Distance to Big Cities**	0.024	0.039	0.054	0.072	0.097	0.115
33	**Gov C Share deflated with GDP prices**	0.021	0.036	0.056	0.075	0.137	0.225
34	**Absolute Lattitude**	0.029	0.033	0.040	0.042	0.059	0.086
35	**Fraction Catholic**	0.019	0.033	0.042	0.056	0.104	0.163
36	**Fertility Rates in 1960s**	0.020	0.031	0.043	0.063	0.108	0.170
37	**European Dummy**	0.020	0.030	0.043	0.049	0.094	0.148
38	**Outward Orientation**	0.019	0.030	0.043	0.054	0.085	0.134

Notes to Table 2: The left hand side variable in all regressions is the growth rate from 1960-1996 across 88 countries. Apart from the final column all statistics come from a random sample of approximately 89 million of the possible regressions including any combination of the 67 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.

Notes to Table 3: The left hand side variable in all regressions is the growth rate from 1960-1996 across 88 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 63 million regressions for $\bar{k} = 5$ to around 44 million for $\bar{k} = 22$.

Notes to Table 4: Entries in the jointness table are calculated by dividing the joint inclusion probability of two variables $x_{i,} x_j$ by the product of the inclusion probability of the two variables individually. If two variables were independent in the posterior (as they are in the prior), this number would be one. Numbers greater than one indicate variables that are help each other in explaining growth (complements), values less than one indicate that variables that explain similar aspects of the dependent variable (substitutes). The index can also be interpreted as measure of collinearity between variables.

Table 4: Jointness Table

#	Variable	1 EAST	2 P60	3 IPRICE1	4 GDPCH60L	5 TROPICAR	6 DENS65C	7 MALFAL66	8 LIFE060	9 CONFUC	10 REVCOUP
1	**EAST**										
2	**P60**	0.97									
3	**IPRICE1**	0.97	1.09								
4	**GDPCH60L**	0.99	1.06	1.13							
5	**TROPICAR**	1.14	1.10	1.09	1.09						
6	**DENS65C**	1.05	1.21	1.19	1.24	1.52					
7	**MALFAL66**	1.15	0.76	0.68	0.76	0.27	0.28				
8	**LIFE060**	1.06	0.51	0.96	1.28	0.67	0.41	1.23			
9	**CONFUC**	0.42	1.06	1.08	0.91	0.40	0.40	0.59	1.03		
10	**REVCOUP**	0.94	0.93	1.01	1.02	0.84	0.77	1.09	1.00	1.34	
11	**LAAM**	0.43	1.13	1.12	1.08	0.20	0.62	0.89	0.82	2.77	1.12
12	**MINING**	0.82	0.88	0.90	1.23	0.77	0.84	1.00	2.03	1.70	1.04
13	**SOCIALIST**	1.05	0.99	0.98	1.07	1.00	1.00	1.17	1.08	0.87	1.50
14	**YRSOPEN**	0.87	0.96	0.81	0.83	0.94	0.65	0.66	0.74	1.50	1.39
15	**MUSLIM00**	0.88	1.09	1.14	1.01	0.52	0.63	0.43	1.74	2.16	1.21
16	**BUDDHA**	0.41	1.08	1.14	0.91	0.85	0.86	0.47	0.69	2.93	1.09
17	**AVELF**	1.10	1.04	1.09	1.08	1.10	1.13	0.72	0.59	0.49	0.87
18	**GVR61**	1.01	0.63	0.87	0.99	0.70	0.47	1.18	1.89	0.99	1.71
19	**DENS60**	0.91	1.12	1.09	1.29	1.12	1.44	0.40	1.00	1.13	1.02
20	**PROT00**	0.91	0.96	0.92	0.54	0.53	0.50	1.80	1.17	1.14	1.07
21	**OTHFRAC**	1.01	0.80	1.06	1.16	0.82	0.47	0.77	2.20	1.03	1.28

		11	12	13	14	15	16	17	18	19	20
	Variable	**LAAM**	**MINING**	**SOCIALIST**	**YRSOPEN**	**MUSLIM00**	**BUDDHA**	**AVELF**	**GVR61**	**DENS60**	**PROT00**
1	**EAST**										
2	**P60**										
3	**IPRICE1**										
4	**GDPCH60L**										
5	**TROPICAR**										
6	**DENS65C**										
7	**MALFAL66**										
8	**LIFE060**										
9	**CONFUC**										
10	**REVCOUP**										
11	**LAAM**										
12	**MINING**	1.61									
13	**SOCIALIST**	0.76	1.02								
14	**YRSOPEN**	0.73	0.98	1.08							
15	**MUSLIM00**	0.52	1.45	0.97	1.32						
16	**BUDDHA**	1.41	1.79	0.88	0.88	2.07					
17	**AVELF**	1.24	0.90	1.05	0.93	0.87	0.50				
18	**GVR61**	1.31	1.70	0.95	0.45	0.78	0.91	0.41			
19	**DENS60**	1.44	0.32	1.12	0.70	0.81	1.02	1.03	1.26		
20	**PROT00**	0.87	1.04	0.94	1.74	0.51	0.62	1.12	1.10	0.81	
21	**OTHFRAC**	1.49	1.84	0.92	0.71	0.58	1.08	0.41	1.95	0.82	0.44