AUSTRALIAN
SCHOOL OF BUSINESS ™

THE UNIVERSITY OF NEW SOUTH WALES

www.business.unsw.edu.au

School of Economics
Australian School of Business
UNSW Sydney NSW 2052 Australia

http://www.economics.unsw.edu.au

Time Varying Determinants of Cross-Country Growth

Rachida Ouysse and Chris Nicholas

School of Economics Discussion Paper: 2008/03

ISSN 1323-8949
ISBN 978 0 7334 2599 8

Time Varying Determinants of Cross-Country Growth

Rachida Ouysse * Chris Nicholas †

February 1, 2008

Abstract

In this paper we investigate the time variation in the short term growth determinants through five subperiod cross sectional growth regressions. We also use a panel regression to analyze the long term pervasive drivers of cross-country growth. A fully *Bayesian* approach in the spirit of Fernandez, Ley and Steel (2001)[4] is used to determine the likely candidates for the *best approximating* growth model. Although the findings support the *optimistic* view that growth regression has empirical merit, there is evidence of time variation of the growth determinants. Some of the variables like, *Ratio of Real Domestic Investment to GDP*, *Real GDP per capita*, and *Sub-Saharian African Dummy* emerge as long term growth indicators, while *Fertility Rate* loses its historical status as "core growth" factor. The findings also show that the convergence effect of *Real GDP per Capita* happens during the middle subperiods. Short run analysis shows evidence of a delay in the growth rate response to the initial conditions followed by an acceleration effect before hitting a plateau.

*Corresponding Author: School of Economics, The University Of New South Wales, Sydney 2052 Australia. Email: rouysse@unsw.edu.au

†This work is part of Nicholas's thesis work to complete requirements for his the *Honours* degree at UNSW.

1 Introduction

Perhaps the most interesting point regarding empirical studies of economic growth is the vast number of studies and conflicting evidence regarding which variables are useful in predicting long term growth. There are many conflicting theories and empirical findings as to what drives economic growth, ranging from persistent technological progress, to investment in human capital, or perhaps prudent government policy, and in many cases any combination of these and other factors. As a result of this, the question became startlingly clear - what drives economic growth?

This entire literature stems from the study conducted by Barro (1991)[1], which yielded empirical support for the theories first proposed by neoclassical growth models such as Solow (1956)[19][1]. Naturally, with so many conflicting theories and results, Levine and Renelt (1992)[11] proposed to run an OLS study with different combinations of variables rather than run a single OLS model with *ad hoc* choice of regressors with simple diagnostic based on the usual t-statistic. The authors argued that it would be beneficial to aggregate the regressions results across the entire model space and investigate which variables seem systematically correlated with economic growth.

The methodology used in this process was that of the extreme bounds test, first proposed by Leamer (1983[10], 1985[10]) where a 2 standard deviation confidence interval was taken on both the upper and lower bounds: if the upper bound was positive and the lower bound was negative, then the variable was labeled non-robust across the model space. The study concluded that virtually no regressors are robust, which led to the initial stance that "nothing can be learned from this empirical growth literature because no variables are robustly correlated with growth"[2]. Of course this view has since been challenged, and more recent empirical studies have supported the more optimistic outcome that some variables are indeed correlated with economic growth and therefore belong to the *"best approximating"* model.

This progression of thought began with Sala-i-Martin (1997)[17] (Sala-i-Martin hereafter), whereby the extreme bounds methodology was rejected as perhaps too harsh a test. This proposition stems from the idea that if enough regressions are run, then any variable can change signs enough times to make a failure of the extreme bounds test somewhat of a forgone conclusion. Sala-i-Martin reinforced this claim under the consideration that "a lot of the variables used in the literature reflect similar economic phenomena so *multicollinearity* among variables is considerable".

As a mean to demonstrate the point, Sala-i-Martin proposed a rather different test applied to the same sort of methodology employed by Levine and Renelt [11]. To begin with, Sala-i-Martin chose 62 variables which have been found to be significant in at least one study previously conducted, 3 of which were included in every regression that was run; *GDP per capita in* 1960, the *Primary school enrolment rate in* 1960 and *Life expectancy in* 1960. Furthermore, the *Average investment rate* was included in most of the studies conducted. The other variables included the *Average growth rate of the population over the period* $1960 - 1992$, the *Average savings rate from* $1960 - 1992$, *Geographical dummy* variables (Sub-Saharan African Dummy, Latin American Dummy) and a measure of the *degree of Capitalism* as the study in Sala-i-Martin

[1]Neoclassical growth models tend to suggest that economic growth is driven by human capital investment and not necessarily technological advancement. These theories also tend to support the notion of income convergence across countries (i.e. poorer countries per capita GDP grows faster than richer countries).

[2]Sala-i-Martin (1997)[17]

was concerned with the time when communism dominated eastern European government. The dependent variable in Sala-i-Martin is the *growth rate of GDP per Capita from* $1960-1992$.

The method Sala-i-Martin[18] employs is to fix the three 'core' variables; *GDP per capita in* 1960*, the primary school enrolment rate in* 1960 and *life expectancy in* 1960 in every regression, and combine them with combinations of the other variables. The inclusion of these three key variables is based on the following criteria:

- GDP per capita in 1960: Included in essentially every regression in the literature and usually found to be significantly negative due to the convergence effect.
- Primary school enrolment rate in 1960: Identified by Barro [1] as an important variable measuring human capital.
- Life expectancy in 1960: Measure of non-educational human capital usually found to be significant.

The *Average investment rate* is also used as a fixed variable by Levine and Renelt [11], and is also widely used in prior studies. Subsequently, Sala-i-Martin[18] conducted studies both including and excluding the *Average investment rate* while taking note of the partial correlation between a certain variable x and growth. The principle is that if investment is included in a regression, and a variable x is still highly correlated with growth, then x is thought to have an effect on growth above the incentive to invest, and there may be a direct correlation between x and growth. Furthermore, if a variable x is correlated with growth when the *Average investment rate* is not included in the regression, then one cannot make any distinction whether or not x has a direct effect on growth, or whether it simply affects growth through incentives to invest. Investment, above all other variables, is singled out in this context due to "the central role that investment plays in growth theory"[3]. Now with the core regressors set, the other variables grouped into sets of three to reflect the point that typically, growth regressions have at least seven right hand side variables. Sala-i-Martin then estimated 4 million regressions using different subsets of explanatory variables. The extreme bounds methods works as follows. Suppose one is interested in whether a variable z is a robust determinant in the growth equation. Let x_j be a set of up to three variables from the pool of $\mathfrak{P}$ of $\mathfrak{N}$ variables previously identified to be related to growth. There are $\mathfrak{M}$ possible combinations of $x_j - sets$ in $\mathfrak{P}$. For each set in $j \in 1, .., \mathfrak{M}$, run the following regression,

$$y = \alpha_j + \beta_{Fj} F + \beta_{z_j} z + \beta_{x_j} x_j + \epsilon_j \tag{1}$$

Where F is a vector of the three fixed variables, z is the variable of interest and x_j is a vector of any one of the combinations of three variables. For each of the $\mathfrak{M}$ regressions, a lower bound $\beta_{z_j} - 2\sigma_{z_j}$ and an upper bound $\beta_{z_j} + 2\sigma_{z_j}$ are computed. A variable is said to be not *robust* if the lower bound is negative and the upper bound is positive.

Sala-i-Martin departed from what he called "extreme test" to using a probability based criterion to evaluate the "robustness" of a variable. Based on the same set of regressions as in the extreme bounds test, Sala-i-Martin constructed a cumulative distribution function across the regressions estimates of β_{z_j}. Because the point estimates in these regressions come from different probability distributions, Sala-i-Martin computed weights proportional to the (integrated)

[3] Sala-i-Martin [18]

likelihood L_j for each of the $\mathfrak{M}$ regressions. An estimate of β_z is then calculated as the weighted average of the point estimates from each regression,

$$\begin{aligned} \widehat{\beta_z} &= \textstyle\sum_{j=1}^{\mathfrak{M}} \omega_{z_j} \widehat{\beta}_{z_j} \\ \omega_{z_j} &= \frac{L_{z_j}}{\sum_{k=1}^{\mathfrak{M}} L_{z_k}} \end{aligned} \tag{2}$$

The reason for this weighting process is that more weight is designated to the models that are closer to the "true" model. Similarly, the variance of each parameter is also calculated as a weighted average;

$$\widehat{\sigma}_z^2 = \sum_{j=1}^{\mathfrak{M}} \omega_{z_j} \widehat{\sigma}_{z_j}^2 \tag{3}$$

Sala-i-Martin uses the probability mass on either side of Zero to compare the relative importance of a variable z to growth. If the distribution of the point estimates $\widehat{\beta}_{zj}$ is Normal across the $\mathfrak{M}$, then for each z, the $CDF(0)$ is computed simply as,

$$CDF(0) = \Phi\left(-\frac{\widehat{\beta}_z}{\widehat{\sigma}_z}\right) \tag{4}$$

where $\Phi()$ is the cumulative distribution function for the Standard Normal.

If the point estimates are not normally distributed across models, Sala-i-Martin calculates a weighted CDF_0^W and non-weighted CDF_0^{NW} for each parameter $\widehat{\beta}_{zj}$. The two averages are computed using the individual

$$\widehat{CDF}(0)_j = Pr\left(\widehat{\beta}_{zj} < 0\right) \tag{5}$$

for $j = 1, .., \mathfrak{M}$ and calculating the quantities,

$$\begin{aligned} CDF_0^W &= \textstyle\sum_{j=1}^{\mathfrak{M}} \omega_{zj} \widehat{CDF}(0)_j \\ CDF_0^{NW} &= \textstyle\sum_{j=1}^{\mathfrak{M}} \frac{\widehat{CDF}(0)_j}{\mathfrak{M}} \end{aligned} \tag{6}$$

The purpose of these two final measures is simply to account for *endogeneity* bias in what may seem more likely models due to the nature of the variables being included. This may be spurious, and the weight placed upon these models may not reflect the "true" fit of the model, therefore, the weighted and non-weighted non-normal CDF's are added as a mean of comparison.

The results of Sala-i-Martin's method give light to a far more optimistic conclusion as to whether or not variables are important in growth regressions. Firstly, by calculating the extreme bounds method on every variable, Sala-i-Martin effectively reproduced Levine and Renelt's[11] results in that the only variable that may be deemed robust of all 62 tested by the extreme bounds test is the fraction of the population that follows the Confucius religion. Now this variable effectively acts as a dummy for the *East-Asian miracle* economies, and serves more as a proxy than as an explanatory variable. Of course, Sala-i-Martin's own statistics tell a far more intriguing story. Of the 59 variables tested, given that 4 have been fixed, it seems that if we take the 95% level of significance, 21 of these variables are significant using the non-normal non-weighted CDF.

Of particular interest, variables outside of the 4 core regressors were found to have CDF's of approximately 1, meaning that they were significant in every regression that they were in. These five variables were the *Number of years that an economy has been open, Fraction of Confucians in a population*, which as mentioned before acts as a proxy for the *East-Asian miracle*, The *Rule of Law, Fraction of Muslims* in a population, which Sala-i-Martin's argues may act as a proxy for oil producing nations and finally a *Political Rights index.* All of these show positive correlation to growth. When the investment rate is not included as a fixed regressor, *Equipment Investment* also charts a CDF equal to 1. The study therefore contradicts the findings of Levine and Renelt[11] and simply concludes that the 'nothing is robust' approach to growth regressions is not necessarily a prudent approach to take.

Subsequent to this study, Fernandez, Ley and Steel (2001b)[6](FLS hereafter) proposed a statistically grounded framework in which Sala-i-Martin's far more optimistic conclusion could be tested. This method is based in *Bayesian* statistics, specifically in the concept of *Bayesian* Model Averaging (BMA), whereby not only parameter uncertainty can be accounted for in a model, but also model uncertainty across the entire model space.

The principle behind this approach is based on the double uncertainty in model specification: theory not only silent about the number of pervasive growth determinant but also their identity. Sala-i-Martin approach is based on certain knowledge that the three core variables are indeed in the "true" model and also certain knowledge about the size of the best model. Sala-i-Martin fixed the model size to 7, limiting the number of regressors that can be used and limiting the possible number of models. FLS[4] approach not only opens up any combination of variables by not fixing any of them, but also lends itself to any model size across all possible models. T

In a sense, *Bayesian* analysis aims to let the data speak, and does not attempt to make deterministic assumptions about what should and should not be included. Through the posterior probabilities, *Bayesian* setting enables a statistically grounded model averaging , unlike Sala-i-Martin (1997), "even though he considers weighing with the integrated likelihood"[4].

The set back to *Bayesian* statistical methods is the provision of a prior probability regarding the researchers beliefs about model and parameter distributions. The implementation of prior probabilities is spurious in the context of growth regressions, due to the fact that all the previous conflicting evidence means that there is no way of knowing which regressors are most likely to be included in the "true" model and just how big that model should be, which leaves us back where we started.

Subsequently, Fernandez, Ley and Steel (2001a)[4] proposed the use of improper non-informative priors in setting up their BMA analysis. Given that the researchers have little or no idea about parameter distributions or model distributions, the prior settings provide very vague information to the model reflecting our own ignorance.

Since the designation of prior parameters is the main concern with *Bayesian* analysis, it was only a natural progression of thought that the *Bayesian* approach taken by Fernandez, Ley and Steel (2001b) would come under a degree of scrutiny. An alternative approach was suggested by Doppelhofer, Miller and Sala-i-Martin (2000)[5], who's *Bayesian* Averaging of Classical Estimates (BACE), used a combination of *Bayesian* statistics and OLS regression theory, in an attempt to utilize the benefits *Bayesian* Model Averaging techniques whilst following in the

[4]FLS[6] also note that Sala-i-Martin gives no insight as to what the integrated likelihood may be, but rather infer it may be the maximized likelihood

"classical spirit of Sala-i-Martin ([17] and [18])".

Doppelhofer, Miller and Sala-i-Martin (2000)[5](DMS hereafter) claimed that since the number of possible regressors K, is large, and the total model space 2^K is extremely large, "fully specifying priors is unfeasible". The criticism here is that by using a purely *Bayesian* framework for estimating a regression[5], the researcher would be forced into arbitrarily choosing prior parameters in a manner which is extremely difficult to interpret. Furthermore, the fact that no sensitivity analysis had been conducted regarding the use of different prior parameters only adds ambiguity to any study conducted in a purely *Bayesian* setting.

In light of this, DMS[5] argued that the use of diffuse priors is a far more appropriate means of specifying no prior beliefs over the model. In a bounded parameter space, the diffuse prior is a simple uniform distribution. In the case of most multiple linear regressions however, the parameter space is unbounded, and as this is the case a uniform distribution cannot be directly assigned. Theoretically, a limit is taken as the prior distribution becomes flat. The use of this prior is justified by DMS[5] in the sense that its imposition generates classical results which yield "posterior distributions identical to the classical sampling distribution of OLS". The major issue with such an approach is that, when dealing with models of different sizes is that when limiting distributions are taken, the mean posterior model size will always be higher than its prior counterpart. Therefore DMS[5] proposed that the limit be taken not as the prior becomes flatter, but rather when the information gained from the data becomes very informative relative to the prior.

FLS[6] proposed a new approach to deciding on what determines the growth rate. The approach is based on *Bayesian* variable selection which integrates out model uncertainty in the posterior distribution of the parameters. Instead of taking an average of point estimates coming from different distributions as in Sala-i-Martin, FLS[4] used the concept of *Bayesian* model Averaging BMA which estimates model parameters by their posterior mean calculated based on the posterior distribution. In general, if θ is a parameter of interest and inference is to be made based on sample data y. *Bayesian* technique start from a prior belief about θ in the form of a prior distribution $\pi(\theta)$. Given the observed data, the posterior distribution of θ is given by Bayes rule,

$$p(\theta|y) \propto p(y|\theta) \times \pi(\theta) \tag{7}$$

A *Bayesian* estimate for θ can be constructed as the value of θ corresponding to the mode of $p(\theta|y)$, it can also be chosen to answer a specific decision theoretic problem based on some Risk function. *Bayesian* model averaging takes the estimate to be the expected value of the posterior mean,

$$\widehat{\theta} = E(\theta|y) \tag{8}$$

In *Bayesian* analysis the whole posterior distribution is readily available and not only the first and second moments. The results of FLS[6] study support the claim of Sala-i-Martin (1997) that certain variables are important in growth regressions and that the initial findings of Levine and Renelt's (1992)[11] extreme bounds study may not be the correct conclusion to draw regarding growth regressions. Firstly, four variables were found to be strongly and robustly related to growth, those being the *Level of income in* 1960, the *fraction of GDP in Mining*, the *Number of Years an economy has been open between 1950 and 1994* and the *fraction of Confucians in the population*, which can be interpreted as a dummy for the *Asian miracle economies* as was the case in Sala-i-Martin (1997). Moreover, another 7 variables were found to be statistically important,

[5]Such as the set up in Fernandez, Ley and Steel [6]

but not to the same degree as the four just mentioned. These are *Life Expectancy in* 1960, the *Primary Schooling Enrolment Rate in* 1960, the *Sub-Saharan African Dummy* Variable, the *Latin American Dummy Variable*, the *fraction of Protestants in a population*, the *fraction of primary exports to total exports and Real exchange rate distortions.*

When compared to Sala-i-Martin's results, it is quite clear that there are some similarities in variables found to be important across both studies. For example, the convergence is consistent in both studies, in that the coefficient on the initial level of income in 1960 is negative, indicating that lower income nations tend to grow faster than richer nations. Moreover, the *fraction of Confucians in a population*, which both studies interpret as a proxy for the *Asian miracle economies*, has a robust and statistically significant and positive effect on economic growth. To take this point even further, the *Number of years that an economy is open* is also positive and statistically significant in both studies.

Of course there are some differences when comparing the two sets of results. Most notably is that DMS[5] found the *fraction of GDP in Mining* has a very high probability of inclusion almost 1 (0.998 in fact), whereas Salai-i-Martin finding was that with a $CDF(0) = 0.817$, the significance of the variable is high, but does not pass the 0.95 standard. Furthermore, the *Latin American and sub-Saharan African* Dummy Variables passes the 0.95 standard in Sala-i-Martin but does not have a posterior probability of inclusion above 0.90 in DMS[5]. In terms of mean model sizes, a sensitivity analysis was also conducted by changing the mean model size from 7 initially to range from 5 to 16. Almost every variable became more likely to be included as the model size became larger (as would be expected!). Amongst these, the most notable was the *primary school enrolment rate in* 1960, which initially had a posterior inclusion probability of 0.627 when the mean model size was 7, but increases to 0.947 as the mean model size is increased to 16. The increase can be seen in almost every other variable, but *primary school enrolment in* 1960 is singled out as showing the greatest sensitivity to mean model size of those variables that had the highest probability of inclusion when the mean model size was 7. Perhaps even more interesting is the fact that as the mean model size is increased, the marginal probability of inclusion of *life expectancy in* 1960 actually gets smaller, albeit only by a small amount. The conclusion drawn from FLS[6] is again in support of Sala-i-Martin's claim that certain variables are indeed statistically important in growth regressions.

By taking on a purely *Bayesian* framework, FLS[4] made no deterministic assumptions about model size or the distribution of parameters across models. Furthermore, with the implementation of non-informative priors, no economic growth theories were imposed to fit the data. In the simple sense, the model that was run effectively allowed for the data to dictate the posterior moments of each parameter and the overall model space. Unlike the model proposed by DMS[5] which assigns independent sampling probabilities on each parameter, FLS[6]) employed the Markov Chain Monte Carlo techniques to provide a purely random draw of models along a Markov Chain. This is an unfortunate necessity of *Bayesian* statistics since all models can't be computed as k gets large, simply because there are 2^k possible models across the entire space. Fernandez, Ley and Steel (2001b) applied BMA methodology to Sala-i-Martin's (1997) data set and derived some interesting results. The measure of the posterior probability of inclusion is similar to Doppelhofer, Miller and Sala-i-Martin [5] in the sense that it is the sum of posterior model probabilities for all the models containing that variable. The measure for a posterior model probability is however different between the two studies. Fernandez, Ley and Steel (2001b) use proper likelihood functions, whereas Doppelhofer, Miller and Sala-i-Martin [5] use *OLS* output of the Sum of Squared Errors (SSE).

Again the *Level of income in 1960*, and the *fraction of Confucians* in a population are over the 90% level of posterior inclusion in a model, which is consistent with the findings of Sala-i-Martin and Doppelhofer, Miller and Sala-i-Martin [5].

Conversely, *Life expectancy in* 1960, also passes the 90% level of posterior inclusion in a model, which supports the findings of Sala-i-Martin, but not DMS, which reports a posterior probability of model inclusion at 0.887 with a mean model size of 7. *Equipment Investment* is the last variable found to be above the 90% level of posterior inclusion in a model, which again supports Sala-i-Martin, but unfortunately is not tested by DMS[5]. Interestingly enough, the *Number of years an economy has been open*, is a little further down the list than is found by both Sala-i-Martin and DMS, presenting a posterior probability of inclusion of 0.502. This is far below the statistic published by Doppelhofer, Miller and Sala-i-Martin [5] which was 0.997, and Sala-i-Martin (1997) with a $CDF(0) = 1$. While these two studies show emphatic support for the inclusion of this variable, it is far more ambiguous in the FLS[6] model.

So while the variables that are statistically significant over each cross-sectional analysis vary, there is some common ground, and the conclusion across each study is the same - there are variables that may be used to explain economic growth.

This analysis has also been extended into a panel data approach by such studies as Islam (1995)[8] and Barro and Sala-i-Martin (1992)[3], by breaking the period under analysis into sub periods. These studies however, were concerned with the notion of convergence as opposed to variable selection. The primary aim of this study is to look at the robustness of growth determinants over time. The contribution to the literature is two fold: (1) investigate time variation in the short term growth determinants by breaking the sample period into five subperiods and performing variable selection over the cross section growth regression. (2) analyzing the likely candidate for long run growth determinant through a panel variable selection. The paper is organized as follows. Section one introduces the statistical methodology. Section two describes the variables and data specifications. Section three presents the empirical results. Finally, section four concludes with comments and discussion.

2 Methodology

Bayesian variable selection methods have become increasingly popular in recent years due to the advances in MCMC computational algorithms. The method used is close in its methodology to the theoretical set up employed by Fernandez, Ley and Steel (2001b). This, as discussed earlier, is the *Bayesian* Model Averaging approach to growth modeling. The approach that this study takes is to split a panel data set into five periods from $1960 - 1984$, and perform five separate cross-sectional studies and determine which variables are consistently important over time. This study will then be compared to a pooled data study which combines all of the five time periods, allowing for a crossover of idiosyncratic periods. As was previously noted, this approach does not fix any variables in any of the regressions that are run, allowing for any combination of the variables to exist as a feasible model. This gives a total of 2^k possible models (Over 33 million models). With one difference from FLS[**?**], is that we will make use of a the MCMC algorithm proposed by Kohn, Smith and Chan (2001)[7] (KSC hereafter). This is motivated by the fact the search space can be very large if we have a large number of potential factors. If there are K candidate regressors, there are 2^K possible combinations. The scheme in

KHS reduces the computational burden by decreasing the visits the algorithm visits to useless subsets. Consider the linear growth regression widely used in the literature,

$$y = \alpha\iota_n + X\beta + \sigma\epsilon \tag{9}$$

where y is a vector of growth rate for n countries, α an intercept also representing mean growth rate, X is an $n \times k$ matrix of growth determinants and β are coefficients representing marginal growth effect of a unit change in each of the variables in X and, ϵ is the error term representing short term deviations from the equilibrium growth equation. Growth theory falls short from giving any guidelines to what factors influencing growth should enter the growth equation in (9). The growth factors in (**??**) are unknown but are assumed to be elements of a finite set $\mathfrak{X}$ of potential variables. Let K be the total number of potential variables in the space spanned by $\mathfrak{X}$. Further, let X^0 be the set of pervasive growth factors. Define the Bernoulli random variable γ_j as: $\gamma_j = \begin{cases} 1 \; if \; X_j \subseteq X^0 \\ 0 \quad otherwise \end{cases}$. Therefore, $\gamma = \{\gamma_j\}_{j=1}^{K}$, is a selector vector over the column of X. Let q_γ be the "Binomial "random variable representing the number of variables in the selected model, therefore

$$q_\gamma = \sum_{i=1,..,K} \gamma_i$$

For a given value of the latent variable γ, define β_γ as the subset of all nonzero elements of β and let X_γ be the collection of those columns corresponding to $\gamma_i = 1$. The latent variable γ therefore defines a model, M_γ, which is defined by

$$\begin{aligned} M_\gamma : y &= \alpha_\gamma\iota_n + \underset{(n\times q_\gamma)}{X_\gamma}\underset{(q_\gamma\times 1)}{\beta_\gamma} + \sigma\epsilon \\ \epsilon &\sim N(0, I_n) \end{aligned} \tag{10}$$

In a compact form, the regressor can be extended to the include the intercept. This will imply a change in the definition of the indicator variable which now will have the first element equal to one with probability one. We adopt this notation for ease of derivation especially for the case of panel regression.

$$y = Z_\gamma\theta_\gamma + \sigma\epsilon \quad \text{where, } Z \equiv (\iota_n, X_\gamma) \quad \text{and, } \theta = \begin{pmatrix} \alpha_\gamma \\ \beta_\gamma \end{pmatrix} \tag{11}$$

The idea behind the *Bayesian* approach is to start from some prior belief about the quantities of interest and to use observed sample information to update these beliefs and to construct posterior probability densities. For a given γ, let $p(\beta, \Sigma|\gamma)$ be the joint prior on the model parameters, the marginal likelihood

$$p(y|\gamma) = \int p(y|\theta_\gamma, \sigma, \gamma)p(\theta_\gamma|\sigma, \gamma)p(\sigma|\gamma)d\theta_\gamma d\sigma \tag{12}$$

Analytical expressions for the full marginal likelihood are possible to obtain for a range of conjugate priors on the model parameters. The aim is to compute the posterior distribution of the latent variable γ. This is readily obtained from the following

$$p(\gamma|y) \propto p(\gamma)p(y|\gamma) \tag{13}$$

where $p(\gamma)$ is the prior on γ. In order to make inferences about γ one needs to be able to compute or at least simulate from the density in (13). This is not analytically feasible and this is where Markov Chain Monte Carlo methods play a crucial role to enable sampling from the full posterior without knowledge of the constant of proportionality in (13).

Fernandez, Ley and Steel (2001b)[6] acknowledges that the choice of prior may have a significant impact on the posterior results ascertained from the model. FLS[4] proposes the use of a "benchmark prior" which is thought to have little influence on posterior results, and would be appropriate to incorporate into analysis where little or no prior information is known . We follow FLS[4] and use non-informative for σ, and a g-prior structure for θ. This corresponds to,

$$\theta|\sigma \sim N\left(\theta_0, \sigma \times \underline{H}_\theta\right) \tag{14}$$

and a prior on σ, $p(\sigma) \equiv \sigma^{-1}$. The covariance matrix $\underline{H}_\theta$ determines the amount of information in the prior and will influence the likelihood covariance structure. In the literature, the specification is simplified to $\underline{H}_\theta = c\underline{V}_\theta$. The preset form $\underline{V}_\theta$, can be chosen to either replicate the correlation structure of the likelihood by setting $\underline{V}_\theta = (Z'Z)^{-1}$, this is also the g-prior recommended by Zellner (1986); or to weaken the covariance in the likelihood by setting, $\underline{V}_\theta = I_K$, which implies that the components of θ are conditionally independent. The scalar c is a tuning parameter controlling the amount of prior information. The larger the value of c, the more diffuse (more flat) is the prior over the region of plausible values of θ. The value of c should be large enough to reduce the prior influence. However, excessively large values can generate a form of the Bartlett-Lindley paradox by putting increasing probability on the null model as $c \to \infty$. In the literature, different values of c were recommended depending on the application at hand. Fernandez, Ley and Steel (2001a) consider many different choices for c, and conclude that an appropriate form is $c = max\{n, k^2\}$. The choice of prior in (2) assumes that there is a common prior for ? across all models. This is a usual practice , and does not seem unreasonable since, by always conditioning on the full set of regressors, X keeps the same meaning (namely the residual standard deviation of y given Z) across models. Now that we have defined priors for the model parameters, we also have to define a prior over the model space to account for model uncertainty (ie. which regressors to include) across the 2^k different models. That is we need to specify the prior distribution of the random variable γ. Each γ_j is a Bernoulli random variable. If we denote by, $P(\gamma_j = 1) = \pi_j$, then

$$p(\gamma) = \prod_{j=1}^{K} \pi_j^{\gamma_j}(1-\pi_j)^{(1-\gamma_j)}$$

For a prior reflecting equal treatment of the factors, $\pi_i = \frac{1}{2}$ and therefore $p(\gamma) = \frac{1}{2^K}$.

Conditional on γ and on the observed data on the growth rate, the full conditional (posterior) probability for the model parameters are given by the following expression(See Ouysse and Kohn (2006)[12] for detailed derivation,

$$\theta|\sigma, y, \gamma, Z \sim N\left(\widetilde{\theta}_\gamma, \frac{c}{1+c}\sigma(Z'Z)^{-1}\right) \tag{15}$$

where $\widetilde{\theta}_\gamma = \frac{c}{1+c}(X'_\gamma X_\gamma)^{-1}X'_\gamma y$.

$$\sigma^2|y, \gamma, Z \sim IG\left(1, \frac{n-1}{2}\right) \tag{16}$$

The next step is to compute the full conditional of the model indicator variable γ. After integrating out uncertainty about θ and σ, the posterior density of γ conditional on y is readily available from the following expression,

$$p(\gamma|y,Z) \propto (1+c)^{-\frac{k_\gamma}{2}} \left(y' \left[I - \frac{c}{1+c} X_\gamma (X'_\gamma X_\gamma)^{-1} X'_\gamma \right] y \right)^{-\frac{(T-1)}{2}} \tag{17}$$

In practice, Markov Chain Monte Carlo MCMC methods are used to explore the posterior distribution of γ and provide a fast and efficient way to identify promising variables witch have high marginal probability of $\gamma_j = 1$. In this study we use the MCMC algorithm proposed by Kohn, Smith and Chan (2001)[7] (KSC hereafter). This is motivated by the fact the search space can be very large if we have a large number of potential factors. If there are K candidate regressors, there are 2^K possible combinations. The scheme in KHS reduces the computational burden by decreasing the algorithm visits to useless subsets. Therefore, it is possible to identify useful $\gamma-$vectors even if one doesn't sweep the whole sample space.

Using Metropolis-Hastings scheme, an ergodic Markov Chain $\gamma^{(0)}, \gamma^{(1)}, ..., \gamma^{(s)}$ of iterates of γ from the full conditional is generated. The sampling distribution of this chain of iterates is used to approximate the distribution $p(\gamma|r)$. Each element $\gamma_k^{(i)}$ in the indicator variable is generated at a fixed or random order from the full conditional distributions. The sampler moves from the current model M_{γ^C} to a new randomly proposed model M_{γ^N} with probability $p = min\left\{1, \frac{p\left(M_{\gamma^C}|y,Z\right)p(\gamma^C)}{p\left(M_{\gamma^N}|y,Z\right)p(\gamma^N)}\right\}$ and remains at M_{γ^C} with probability $1/p$. The KSC [7] algorithm minimizes visits to less likely models and only models with high posterior probabilities will be covered, yet this still allows us to visit a decent proportion of the total model space.

3 Specifications and Data

The data set used in this study is a panel data set from 1960-1990, covering national statistics from 138 countries. This data set is taken from Barro and Lee (1994)[2] -" Data Set for a Panel of 138 Countries" which is accumulated data from several other studies. The panel is broken up into 6 sub-periods, $1960-64$, $1965-69$, $1970-74$, $1975-79$, $1980-84$ and $1985-1990$. Values are given as either averages over each five year period or as a value corresponding to the beginning of each period . The analysis consists of 25 variables, all of which are measured over the five sub-periods over the course of $1960-84$. The sub-period $1985-1990$ was excluded from the analysis because most of the variables in this data set were not measured over that period. Furthermore, certain variables were excluded from the analysis as they were not measured in either the first period, or in some cases, the first two periods , whilst other variables were taken as averages over the period $1960-84$, these were also excluded as they give no real insight into estimating from a panel data set. Overall, the variables that were ultimately included are:

- Pupil/Teacher Ratio in Primary School: Used as a proxy for the quality of education provided in primary school. This works under the assumption that the fewer students per teacher, the better the degree of education.

- Pupil/Teacher Ratio in Secondary School: Used as a proxy for the quality of education provided in primary school. Again this works under the same assumption as above.
- Ratio of Exports to GDP: This does face the issue of endogeneity, but is included nonetheless.
- Ratio of Imports to GDP: Like the ratio of exports to GDP, this also faces an endogeneity problem, but is also included nonetheless.
- Exchange Rate (First Year) " Gross Secondary School Enrolment Ratio (First Year): Used as a proxy for investment in human capital.
- Gross Higher School Enrolment Ratio (First Year): Implemented by Fernandez, Ley and Steel (2001b). Reflects the tendency of developed nations to move towards tertiary industry.
- Gross Primary School Enrolment Ratio (First Year): Used as a proxy for investment in human capital. First suggested by Barro (1991).
- Sub-Saharan African Dummy
- OECD Dummy
- Latin American Dummy
- East-Asian Dummy: Fraction of Confucians in a population is thought to be a proxy for this variable in previous studies .
- Ratio of Liquid Liabilities to GDP: Shows the ability of a nation to cover its short term liabilities. Again this suffers from an endogeneity problem but is included in the analysis nonetheless.
- Black Market Premium: The difference between the actual exchange rate and the premium payed for black market transactions. Calculated as the (Black Market Exchange Rate/Official Exchange Rate) -1
- Terms of Trade Shocks: Growth rate of export prices - growth rate of import prices.
- Fertility Rate: Children per woman
- Infant Mortality Rate
- Life Expectancy: Life expectancy at age 0, taken as an average over the period.
- Growth Rate of Population: Taken as an average over the period.
- Real GDP Per Capita (First Year): Taken as the value at the beginning of the period.
- Ratio of Real Domestic Investment to GDP: Again this suffers from an endogeneity problem but is included in the analysis.
- Number of Revolutions per year: Taken as an average over the period
- Ratio of Total Workers to Population: Taken as an average over the period.
- Area (millions km^2)

- Distance (to capitals of world's 20 major exporters)

As is the case with virtually every data set, there were many missing observations. As a result, averages were taken over each period and imputed in missing data points. It was considered to simply ignore those observations with missing values, but as this tended to change over time, for the sake of a consistent panel-type analysis, means were imputed to complete the data set, to ensure that the same variables were being studied across every period. The data was then pooled for further analysis.

4 Posterior Results

The results above are based on the MCMC sampler described in the method section of this paper using a burn-in value of 1 million draws and a subsequent 2 million drawings. This is designed so that any information from the prior assigned before running the model, no matter how vague, can be absorbed by the model and the data is able to dictate the posterior moments of the variable and model space. The period was initially broken up into five sub-periods to analyze which variables persistently present some explanatory power on economic growth. The pooled data analysis was then conducted to account, in some part, for any inter-temporal effects that may have been missed in the periodical analysis.

Inclusion probabilities are measured against the standard set by Fernandez, Ley and Steel[6], of above 90% for a high probability of inclusion and above 10% and below 90% for a less convincing but not completely disregard able probability. Figures below 10% give little support for including a variable in the *best approximating model.* A key advantage of the set-up employed by FLS[6] is that the study accounts for *jointness* in variables. Simply put, say that two variables x_1 and x_2 were highly collinear, (i.e. they both tend to estimate the same effect), then these two are disjointed. Similarly, two regressors that tend to add posterior probability to a model are considered jointed. Having this built into the model helps circumvent the issue of multi-collinearity, which was one of the major issues faced by the study of Sala-i-Martin (1997). It is also noted by FLS[6] that confidence intervals in posterior estimates tend to be wider than classical *OLS* estimates. This is interpreted in the sense that classical estimates may underestimate uncertainty.

4.1 Variable Summaries

1. Pupil/Teacher Ratio in Primary School : Posterior results show virtually no indication of importance in any of the 5 sub periods, with a posterior inclusion probability of less than 7% in all periods. The pooled analysis reinforces this point, showing a posterior inclusion probability of 3.89%. The posterior moments of this distribution also indicate a great deal of uncertainty in each sub-period and the pooled analysis, again indicating a negligible degree of significance, thus indicating that this variable has essentially no impact on economic growth.

2. Pupil/Teacher Ratio in Secondary School : Like its primary school counterpart, there is very little evidence supporting the inclusion of this variable in a growth regression based on the data used in the periodical analysis. Over the five year sub-periods, this

Table 1: *Bayesian* Model Averaging BMA Probability of Inclusion. Figures in bold are posterior probability of inclusion > 10%, probabilities are in %

Variables	1960-64	1965-69	1970-74	1975-79	1980-85	Pooled
Pupil to Teacher Ratio in Primary School	3.66	5.55	4.44	6.76	4.32	3.89
Pupil to Teacher Ratio in Secondary School	4.22	5.25	6.20	3.88	5.07	**12.77**
Ratio of Exports to GDP	3.92	3.83	5.50	4.84	3.60	5.48
Ratio of Imports to GDP	4.35	6.28	3.74	**10.84**	3.46	5.67
Exchange Rate (First Year)	3.79	**19.04**	4.52	3.75	3.46	4.12
Terms of Trade Shocks	4.11	4.66	**46.15**	3.61	3.91	**15.18**
Black Market Premium	3.91	4.40	7.71	**99.98**	**11.43**	**100**
Ratio of Liquid Liabilities to GDP	4.47	**24.03**	3.45	6.74	3.65	3.84
East Asian Dummy	5.20	**17.88**	8.24	6.22	3.73	3.80
Latin American Dummy	8.09	4.99	3.68	**39.11**	**99.98**	**99.84**
OECD Dummy	6.77	5.91	**11.44**	4.37	6.28	7.59
Sub-Saharian African Dummy	**59.20**	4.11	**11.52**	3.98	**46.89**	**100**
Gross Higher School Enrolment Ratio (First Year)	4.45	6.40	8.39	3.63	3.72	6
Gross Primary School Enrolment Ratio (First Year)	**20.36**	**33.70**	5.69	**13.36**	5.45	4.48
Gross Secondary School Enrolment Ratio (First Year)	**10.21**	4.78	4.90	8.19	**25.04**	**42.98**
Fertility Rate	5.93	8.47	5.48	3.95	**69.61**	**64.22**
Infant Mortality Rate	5.75	8.13	5.17	6.12	4.73	5.23
Life Expectancy	6.07	**10.38**	4.43	**99.75**	**15.79**	**18.56**
Growth Rate of Population	4.09	3.67	7.84	**14.48**	**13.28**	**17.22**
Real GDP per Capita (First Year)	4.99	**88.63**	**64.64**	**99.65**	**15.66**	**100**
Ratio of Real Domestic Investment to GDP	**51.05**	**99.40**	**99.89**	3.61	**10.89**	**100**
Number of Revolutions per year	3.81	6.41	3.61	5.24	8.92	6.75
Ratio of Total Workers to population	4.46	**12.62**	5.62	3.56	7.75	8.53
Area (millions km^2)	6.03	4.34	3.72	3.95	4.29	6.49
Distance (to capitals of world's 20 major exporters)	**38.74**	3.29	3.62	4.47	3.41	7.20

Table 2: Unconditional Posterior Moments from *Bayesian* Analysis

Variable	**1960-64**	**1965-69**	**1970-74**	**1975-80**	**1980-85**	**Pooled**
Pupil/Teacher Ratio in Primary School	0.00012 (0.00467)	0.00119 (0.00691)	-0.00086 (0.00766)	-0.00216 (0.01062)	0.00081 0.00675)	0.00014 0.00256)
Pupil/Teacher Ratio in Secondary School	0.00095 (0.00949)	-0.00157 (0.00938)	0.00344 (0.01818)	-0.00079 (0.00953)	-0.00158 (0.01000)	-0.00365 (0.01144)
Ratio of Exports to GDP	0.02373 (0.43270)	0.00851 (0.60840)	-0.12120 (0.84450)	-0.05721 (0.75570)	-0.01814 (0.33990)	0.04094 (0.27320)
Ratio of Imports to GDP	0.05171 (0.52700)	0.17050 (0.94340)	0.04052 (0.70490)	0.34580 (1.27800)	0.00320 (0.34220)	0.04826 (0.30030)
Exchange Rate (First Year)	0.00002 (0.00054)	0.00083 (0.00197)	-0.00009 (0.00066)	0.00004 (0.00056)	0.00002 (0.00025)	-0.00002 (0.00021)
Terms of Trade Shocks	-0.20540 (2.10300)	-0.20070 (1.37600)	3.04700 (3.75800)	0.03522 (0.88800)	0.25330 (2.53600)	0.52130 (1.45400)
Black Market Premium	0.00679 (0.10270)	-0.01344 (0.10490)	-0.04388 (0.19390)	-1.15200 (0.24350)	-0.02893 (0.09510)	-0.67670 (0.12390)
Ratio of Liquid Liabilities to GDP	0.03422 (0.36100)	0.64100 (1.29500)	-0.00651 (0.27810)	0.11260 (0.56640)	0.01889 (0.26800)	-0.00619 (0.13800)
East-Asian Dummy	0.03139 (0.22940)	0.23890 (0.58760)	0.09422 (0.39620)	0.05973 (0.31900)	0.01148 (0.18540)	-0.00413 (0.08470)
Latin American Dummy	-0.06194 (0.28010)	-0.02444 (0.17170)	-0.00934 (0.14000)	-0.60780 (0.86250)	-3.07800 (0.67980)	-1.39000 (0.32430)
OECD Dummy	0.05433 (0.31070)	-0.05347 (0.36920)	-0.16320 (0.61180)	-0.03217 (0.35810)	-0.06097 (0.33400)	0.04804 (0.22200)
Sub-Saharan African Dummy	-0.94070 (0.90860)	-0.01328 (0.15130)	-0.12170 (0.40580)	-0.01344 (0.18810)	-0.83030 (1.01000)	-1.92500 (0.34420)

Table 3: Unconditional Posterior Moments from *Bayesian* Analysis-Continued

Variable	1960-64	1965-69	1970-74	1975-80	1980-85	Pooled
Gross Higher School Enrolment Ratio (First Year)	0.14920 (1.67000)	0.36910 (1.99900)	0.54640 (2.54700)	-0.01597 (0.93300)	-0.01645 (0.70190)	-0.12310 (0.76320)
Gross Primary School Enrolment Ratio (First Year)	0.41660 (0.95380)	0.75590 (1.20400)	0.07161 (0.41740)	-0.35290 (1.06300)	-0.08655 (0.53590)	0.01268 (0.15790)
Gross Secondary School Enrolment Ratio (First Year)	0.25430 (0.95280)	0.05229 (0.51520)	0.01919 (0.51340)	-0.24430 (1.04200)	-0.85020 (1.69200)	-0.96780 (1.27600)
Fertility Rate	-0.00993 (0.06338)	-0.02323 (0.09729)	-0.00646 (0.07594)	0.00473 (0.06572)	-0.44080 (0.35600)	-0.21120 (0.18530)
Infant Mortality Rate	-0.26720 (2.09000)	-0.74420 (3.35100)	0.35850 (2.84600)	0.99930 (6.64300)	-0.03967 (3.13700)	0.05982 (1.65000)
Life Expectancy	0.00091 (0.01218)	0.00560 (0.02087)	-0.00014 (0.01200)	0.24120 (0.05719)	0.01340 (0.03667)	0.00939 (0.02370)
Growth Rate of Population	-0.26670 (4.91500)	-0.29930 (5.09500)	1.92500 (8.58800)	-4.82800 (13.75000)	-6.15200 (19.83000)	-3.47800 (8.96400)
Real GDP Per Capita (First Year)	0.00000 (0.00004)	0.00011 (0.02488)	-0.00021 (0.00019)	-0.00061 (0.00014)	-0.00003 (0.00008)	-0.00039 (0.00007)
Ratio of Real Domestic Investment to GDP	4.04000 (4.52900)	14.46000 (3.82400)	14.62000 (3.56400)	0.02711 (0.77940)	0.68670 (2.37000)	10.13000 (1.73500)
Number of Revolutions per year	0.00148 (0.17170)	0.08455 (0.42160)	-0.01229 (0.20040)	-0.04428 (0.27910)	-0.11890 (0.46380)	-0.03345 (0.16920)
Ratio of Total Workers to Population	0.04036 (0.66940)	0.51230 (1.59200)	-0.16240 (0.95620)	0.02014 (0.67370)	-0.32570 (1.55400)	0.17530 (0.75020)
Area (millions km^2)	0.00009 (0.02736)	0.00000 (0.00004)	0.00000 (0.00004)	0.00000 (0.00004)	0.00001 (0.00004)	0.00001 (0.00003)
Distance (to capitals of world's 20 major exporters)	-0.11830 (0.16940)	-0.00045 (0.02417)	-0.00080 (0.03116)	-0.00493 (0.04001)	-0.00099 (0.02731)	-0.00577 (0.02768)

variable does not pass the 10% standard for model inclusion making it not even marginal in light of shorter term modeling. However, its inclusion probability in the pooled analysis is 12.77%, indicating that over a greater period of time, the ratio of secondary school students to teachers may be a significant variable in a growth regression, although this cannot be concluded with any real confidence given the large standard error of the its slope coefficient relative to its mean value from its posterior moments. Nonetheless, based on the results of this analysis, it seems more likely candidate for inclusion in the "true" model than its primary school counterpart.

3. Ratio of Exports to GDP: The ratio of exports to GDP was initially thought to suffer from some *endogeneity* bias due to the fact that it's taken as an average over each 5 year sub-period. This does not seem to be an issue however, as the posterior inclusion probability for this variable is never higher that 5.50% in any of the five sub-periods or the pooled analysis, demonstrating no persistence in explaining growth across the entire period. This is supported by the high standard errors relative to the point estimate of its slope coefficient in each analysis.

4. Ratio of Imports to GDP: The results for this variable tell a very similar story to the ratio of exports to GDP. There is little evidence supporting its inclusion in the "true" as its inclusion in any model doesn't seem to add any explanatory power, reflected by the low posterior inclusion probabilities from *Table 1.* There is one exception to this however, in the period $1975-80$, there seems to be slight support for the inclusion of this variable over this time period, with an inclusion probability of 10.84%, however, much like the figures for pupil/teacher ratio in secondary school, this cannot be said with any real confidence given the large standard errors in each period, including the period $1975-80$.

5. Exchange Rate (First Year): Exchange rates in the first year of each sub-period seem to have very little bearing on economic growth either. Even though it shows a model inclusion probability of 19.04% in $1965-69$, it does not show any inclusion probability of higher than 5% in any one of the other analysis, including the pooled study. The same conclusion can be drawn for this variable as for the *ratio of Imports to GDP*, that its inclusion in the true model is highly unlikely, given the low inclusion probabilities and high standard errors in each period.

6. Terms of Trade Shocks: Terms of trade shocks show more promise in terms posterior inclusion, given that it has an inclusion probability of 46.15% in $1970-74$ and an inclusion probability in the pooled regression is 15.1%. This seems more likely to be included in the "true" model, and shows at least some promise in a short term regression, although it is shows no persistence in explaining growth over all the sub-periods. Furthermore, its inclusion in a longer period study is also a possibility, although not an exceptionally high one, and seems the most likely candidate of all the variables discussed thus far, yet it still suffers from persistently wide confidence intervals in each analysis.

7. Black Market Premium: This is the first of five variables with a posterior model inclusion of higher than 90% in either the pooled or sub-period analysis. In fact, its inclusion in the pooled regression and the period $1975-80$ seem somewhat of a forgone conclusion with inclusion probabilities of 100% and 99.98% respectively. Of more interest perhaps, is the fact that this variable did not seem a likely candidate in a growth regression in the first three periods, but became very likely in the period $1975-80$ and to a lesser extent $1980-85$. This gives rise to the notion that while a variable may not be important over

each period in question, a high inclusion probability in one ore two of the sub-periods may indicate its use in a long term growth regression, which is reflected by the high pooled analysis inclusion probability whereby the idiosyncratic periods are allowed to interact with each other. Its coefficient is also negative in 5 of the six analyzes, indicating that a higher premium on black market prices seems to correspond with lower levels of economic growth, purporting the importance of a black market, which seems contradictory to most growth literature which supports the rule of law. This variable has a much lower inclusion probability in FLS[6] of 0.157.

8. Ratio of Liquid Liabilities to GDP: Again while this variable may seem more significant due to an *endogeneity* bias, its low inclusion probabilities over four of the five sub-periods and the pooled data analysis indicate that this variable has very little explanatory power on economic growth. The only period that showed some promise for this variable was $1965 - 69$ with an inclusion probability of 24.03%. However, the results of the other periods coupled with the large standard errors in every period indicate that the inclusion of this variable in the "true" model is marginal, much like the other endogenous variable ratio of imports to GDP.

9. East-Asian Dummy: This is perhaps the most interesting statistic of the entire study. While this variable thought to be very important from previous studies it appears from the statistics in this analysis that this variable shows very little significance in growth regressions. In only one of the 6 analyzes does the inclusion probability cross the 10% standard, which is in $1965 - 69$. With no signs of persistence across time or in the pooled analysis, the lack of significance in this variable may mean that the fraction of Confucians in a population may stand as its own explanatory variable and not necessarily as a proxy for this dummy variable. This point is only further reinforced by the large standard errors, giving no certainty about the significance of its slope coefficient.

10. Latin American Dummy: This is the second of five regressors that appear highly likely candidates for inclusion in the true model in the pooled regression and follows in the same vein as the Black Market Premium, in that its inclusion probability spikes in the last two sub-periods. This gives further support to the claim that a variable need only be significant in one or two sub-periods to be significant when the idiosyncratic periods are allowed to interact, such as in the pooled analysis. As can be seen in table 2, the standard error is small relative to the coefficient in both the sub-period 1980-84 and the pooled data analysis, indicating some certainty as to the significance of this variable in growth regressions. This is in conjunction with the findings of previous studies .

11. OECD Dummy: Whether or not a country is a member of the OECD seems to have relatively no impact on explaining economic growth. Low inclusion probabilities across each analysis show very little support for adding this variable into the "true" model, with only one inclusion probability just edging over the 10% standard in 1970-74. This, coupled with large standard errors relative to the estimate of the slope coefficient, shows very evidence to support the significance of this variable in a growth regression.

12. Sub-Saharan African Dummy: This is the third of five variables which seem to be highly correlated with long term growth, with an inclusion probability in the pooled analysis of 100%. This variable is also the first variable discussed thus far that shows some signs of significance in 3 of the 5 sub-periods. This dummy variable has been sighted as very important in prior studies , but in terms of *Bayesian* analysis, this is the first inclusion

probability that passes the 90% standard in a long term growth model. This figure is reinforced by the small standard error relative to the slope estimate in the pooled data analysis. This differs from the other highly significant variables in the pooled analysis in that it is the only one that does not pass the 90% standard in any sub-period, but its marginal significance in 3 of the 5 periods seem to make it a suitable candidate for consideration in the "true" model.

13. Gross Higher School Enrolment Ratio (First Year): This variable is a measure of enrolment rates in tertiary education in a country. Interestingly enough, this shows very little signs of significance in any one of the 5 sub-periods, which is a rather interesting statistic given the trend of developed economies to transfer human capital away from primary and secondary industry into tertiary industry. Furthermore, there seem to be no long term effects from the pooled analysis, which is consistent with the large standard errors from the posterior moments in table 2. This is consistent with the results found by Fernandez, Ley and Steel (2001b) and Salai-i-Matrin (1997).

14. Gross Primary School Enrolment Ratio (First Year): This variable was initially cited by Barro (1991) as a proxy for human capital investment in education. Sala-i-Martin (1997) notes that the results regarding this variable have been mixed from previous studies. This study is no exception. While this, like the Sub-Saharan African Dummy is only one of a few variables that pass the 10% standard for model inclusion in 3 of the 5 sub-periods. However, while it shows some persistence in the sub-period analysis, the pooled data regression shows very little support for including this variable in the "true" model. The ambiguity surrounding the inclusion of this variable is only made worse by the large standard errors pertaining to this variable. Therefore, it is still unclear whether or not this variable could be included in a growth regression.

15. Gross Secondary School Enrolment Ratio (First Year): The inclusion of secondary school enrolment seems to have some weight in the pooled data analysis, with an inclusion probability of 42.98%. Unlike its primary school counterpart, the secondary school enrolment rate shows some signs of significance in only 2 of the 5 sub-periods and funnily enough, has no inclusion probability higher than the highest inclusion probability for gross primary school enrolment. This suggests that there may be some evidence of inter-temporal effects between idiosyncratic periods. While it may not have a great deal of use in short term regressions, the pooled analysis gives some weight to including this variable in a longer term model. These results give more support to the inclusion of this variable than previous studies .

16. Fertility Rate: The fertility rate shows some signs of that it may be included in the true model. While it passes the 10% standard in only one of the five sub-periods analyzed, the inclusion of this variable in the pooled data analysis is better than marginal with a probability in inclusion on 64.22%. This again seems to follow the direction f the Black Market Premium or the Sub-Saharan Dummy in the sense that a variable may need only be significant in one or two sub-periods to be significant in an inter-temporal setting. The significance of this variable may also be a lagged effect, given that high fertility rates will add to the labor force years after birth, this effect would have been captured by allowing the sub-periods to interact in the pooled data analysis. No degree of confidence can be given from its posterior moments however, as it appears that only variables with inclusion probabilities higher than 90% have narrow confidence intervals, giving no further insight as to the significance of this variable.

17. Infant Mortality Rate: This variable shows no signs in any of the 5 sub-periods or the pooled data analysis for inclusion in the true model. This is another variable thought to have some degree of endogeneity bias, in that countries may experience high infant mortality rates as poor GDP per capita growth leaving rendering consumers unable to afford proper infantile care. Its posterior moments show a mixed effect, in that its slope coefficient is negative in some cases and positive in others, but seeing as there are very high standard errors on all the estimates of its posterior moments, little can be inferred about the effect that this has on economic growth.

18. Life Expectancy: Life expectancy is one of the three core regressors cited by Sala-i-Martin (1997). FLS[6] also note a high probability of model inclusion of 0.946, which is only one of four variables to pass the 90% standard. The results from this study however, do not assign the same degree of confidence in including this variable in the "true" model from the pooled data analysis with an inclusion probability of 18.56%, even though it shows persistent significance in the 5 sub-periods, passing the 10% standard in two and even the 90% standard in the period $1975 - 79$. This seems to go against the pattern in the variables Black Market Premium and the Sub-Saharan Dummy, and tends to follow in the same vein as the Primary school enrolment rate. On average the effects of this variable seem positive, which is its effect in the only period where the standard error is low relative to the slope estimate, which is 1975-1979. Only in the period 1970-74 is the effect negative, although this effect is accompanied by a rather high standard error.

19. Growth Rate of Population: Population growth shows a slightly smaller inclusion probability as Life Expectancy in the pooled data analysis, with an inclusion probability of 17.22%. This is interesting as its sub-period figures demonstrate far lower significance in this variable as opposed to *Life Expectancy*, passing the 10% standard in only two periods, indicating that there may be a time effect unaccounted for in the sub-period analysis prevalent in estimating this variable. The significance of his variable, like the Black Market Premium and the Sub-Saharan Dummy seems to elevate in the last two periods only.

20. Real GDP Per Capita (First Year): This is the fourth of five variables that pass the 90% standard for model inclusion in the pooled data analysis. Its also one of only two variables that pass the 10% standard for model inclusion in four of the five sub-periods. Its inclusion in variable selection has been a standard since Barro (1991) showed empirical evidence for the neoclassical concept of income convergence (i.e. Economic growth is negatively related to the initial value of income, so that poorer nations tend to grow faster than richer ones). Its inclusion probability in FLS[6] is equal to 1, indicating that it is essentially a forgone conclusion to include this variable in a long-term growth model. It is also cited as one of the core variables used in Sala-i-Martin's (1997) analysis. The evidence from this set of results indicates emphatic support for the conclusions drawn by these earlier studies, with a pooled data analysis inclusion probability of 100%. Naturally, there will be some *endogeneity* bias in this figure, due to the fact that the data has been pooled and GDP in each period will be a function of GDP at the beginning of the previous period and the average growth rate of GDP over the previous period. Furthermore, the fact that it passes the 10% standard in 3 of the 5 sub-periods and the 90% standard in another indicates its explanatory in shorter term modeling of growth as well. On a side note, the posterior moments from table two indicate a predominantly negative effect on growth, which is consistent with the concept of income convergence. However, an inspection of the loadings of this variable shows that for the first two subperiods, $1960 - 64$ and $1965 - 69$, the

coefficient was first zero and then positive and insignificant (delay effect). In the middle two sub periods, this coefficient becomes larger and negative (acceleration effect) before declining in the last period (plateau effect).

21. Ratio of Real Domestic Investment to GDP: This is the final variable to pass the 90% standard in the pooled data analysis. Furthermore, it is the second of two variables which passes the 10% standard in four of the five sub-periods, crossing the 90% standard in two of these periods. The investment rate is considered the fourth core variable in Sala-i-Martin (1997), and is tested in his subsequent analysis in that study. The high inclusion probability of equipment investment (0.942) and the middle range posterior probability of non-equipment investment (0.431) published by Fernandez, Ley and Steel (2001b) also indicates that investment is another very important variable to consider when modeling economic growth. Like Real GDP Per Capita, its persistent inclusion probability over the four sub-periods indicates its importance in short term modeling as well. Again, in the pooled regression, there may be some inter-temporal effects which contribute to the very high (100%) posterior inclusion probability, which is indicative of its use in long-term growth regressions as well. Positive Slope coefficients indicate that there is a positive relationship between investment growth as a fraction GDP and actual GDP growth.

22. Number of Revolutions per year: Posterior results show virtually no indication of importance in any of the 5 sub periods, with a posterior inclusion probability of less than 9% in all periods. The pooled analysis reinforces this point, showing a posterior inclusion probability of 6.75%. The posterior moments of this distribution also indicate a great deal of uncertainty in each sub-period and the pooled analysis, again indicating a negligible degree of significance, thus indicating that this variable has essentially no impact on economic growth.

23. Ratio of Total Workers to Population: Results show virtually no indication of importance in any of the 5 sub periods, with the exception of the period $1965-69$ where it passes the 10% standard - the only time across the entire analysis. The pooled analysis reinforces this point, showing a posterior inclusion probability of 8.53%. Indicating a negligible degree of significance, and no real evidence supporting its inclusion in the "true" model. This is consistent with the results of Sala-i-Martin (1997b) and Fernandez, Ley and Steel (2001b).

24. Area (millions km^2): The geographic size of a country again has very little bearing on economic growth in any one of the sub-periods or the pooled data analysis. This again is consistent with the results of Sala-i-Martin (1997b) and Fernandez, Ley and Steel (2001b).

25. Distance (to capitals of world's 20 major exporters):Distance to the worlds major exporters makes a noticeable contribution to explaining growth in the period 1960-64 with a posterior probability of inclusion of 38.74%. After this period however, the inclusion probability is never above 5%. This is consistent with the inclusion probability of the pooled data analysis of 7.20%. High standard errors from the posterior moments in table 2 make any inference from the sign of the slope coefficient unfeasible.

4.2 Model and Period Summaries

The results in *Table 4* are the posterior model space moments given a uniform prior over the model distribution. The results are fairly clear cut, indicating that over a shorter time frame,

Table 4: *Bayesian* Model Averaging Posterior Probability

variable	1960 – 64	1965 – 69	1970 – 74	1975 – 80	1980 – 1985	Pooled
Average model size	2.72	3.96	3.40	4.64	3.85	7.50
Standard deviation of model size	1.00	1.22	1.15	1.12	1.18	1.24

the average model size is far smaller. When all the data is pooled, the average model size is significantly larger than any of the five sub-periods. This can be a result of the greater degree of variation in the dependent variable requiring a greater degree of explanation that can only be achieved if the number of explanatory variables grows with time and therefore variation. Over the entire time period, the results seem to coincide with Salai-i-Martin's (1997) conclusion that the mean model size constitutes approximately 7 variables. One of the major advantages of the set-up established by FLS[6] is that with a formalized model averaging process accounting for *jointness* amongst variables, the most likely models can be determined, hence avoiding the issue of *multicolinearity* that is thought to be prevalent in Sala-i-Martin's (1997) approach. This is a result of the MCMC method employed in the analysis, in that it is designed to only move between models with a high posterior mass in order to capture the greatest posterior mass of the entire model space without having to filter through every model which is computationally unfeasible as noted in the method section. The most likely models in each period are listed below. Bearing

Table 5: Most Likely models for Growth over the subperiod and for the pooled sample

Period	Variables in M_γ	Posterior probability $p(\gamma\|y)$
1960 – 64	Sub-Saharian African Dummy Ratio of Real Domestic Investment to GDP	7.56%
1965-69	Real GDP per Capita Ratio of Real Domestic Investment to GDP	8.78%
1970-74	Terms of Trade Shocks Real GDP per Capita Ratio of Real Domestic Investment to GDP	8.53%
1975-79	Black Market Premium Life Expectancy Real GDP per Capita	15.47%
1980-84	Latin American Dummy Fertility Rate	7.91%
Pooled	Black Market Premium Latin American Dummy Sub-Saharian African Dummy Secondary School Enrolment rate Fertility Rate Real GDP per Capita Ratio of Real Domestic Investment to GDP	9.88%

in mind that posterior model probability is calculated by equation (17), it is quite clear that the most likely model in each period and the pooled analysis comprises of variables that show a high posterior probability of inclusion in each respective period. Take, for example, the model for the period 1960 – 64. The most likely model is a function of the two variables Sub-Saharan African Dummy and Ratio of Real Domestic Investment to GDP, both of which happen to have the highest posterior probability of inclusion in that periods, both scoring 59.20% and 51.05% respectively. The same can be said for every other period in the analysis.

Of more interest is the model that corresponds to the pooled data analysis. Allowing for an additional dimension (time) can shed some light on the robustness of growth factors over time. The most likely model in this analysis has seven explanatory variables, which coincides with the mean model size noted by Sala-i-Martin (1997) and is in the range of $6 - 12$ variables corresponding to Fernandez, Ley and Steel (2001b)[6] 75 most likely models. It just so happens that this most likely model is made up of the seven variables with the highest posterior probability of inclusion of the 25 that were tested, which again indicates that there is some credit to the method implemented in this study. Further insight can be gained from these statistics; namely that number of regressors found significant appears positively related with the time period studied, or more to the point, the length of time. This is an expected result as a greater period of time inherently comes with greater variability in the dependent variable, and naturally a greater possibility of foreseeable economic forces that can affect it.

5 Conclusion

The analysis conducted indicates that there is merit in the growth regression and agrees with this "optimistic" stance adopted by Sala-i-Martin (1997), Fernandez, Ley and Steel (2001b)[6] and Doppelhofer, Miller and Sala-i-Martin [5], that certain variables are useful in modeling economic growth. Using *Bayesian* variable selection this study concludes that certain variables are important regressors in modeling economic growth. We do however, disagree on what those variables are based on persistence in inclusion probabilities over five sub-periods and a pooled data analysis.

Of the 25 variables tested, not one single regressor maintains a significant probability of inclusion over every one of the five sub-periods. It is quite clear though that over time the most consistent short term factors were real GDP per capita (First Year) and the Ratio of Real Domestic Investment to GDP. Whilst neither of these two variables passes the 90% standard for inclusion, they are the only two variables which pass the less convincing 10% in four of the five sub-periods, making them the two most consistently significant variables in the analysis. Bear in mind that the posterior probability of inclusion is equal to the sum of the posterior model probabilities for all the models containing that variable. The pooled analysis however, may give a little more insight as the variation in both the dependent and independent variables are allowed greater variation when the entire 25 year period is considered in the one model.

The findings indicate that the inclusion of *GDP per capita (First Year)*, *Ratio of Real Domestic Investment to GDP* , *Sub-Saharan African Dummy*, the *Black Market Premium* and the *Latin-American Dummy* in the **best approximating** model seem almost certain with inclusion probabilities very close (*Latin American Dummy* with 99.84%) or equal to 100%.

The statistical properties of the parameters estimates reinforce these high inclusion probabilities in each analysis. These variables work well in conjunction with one another, eliminating any issue of collinearity, as they make up five of the seven variables comprising the model with the greatest posterior probability in the panel data analysis.

The same applies to the sub-periodical analysis, as the models with the highest posterior probability all consist of the only those variables with the highest posterior inclusion probability in each respective period. Interestingly enough though, the *East-Asian Dummy variable* has very low inclusion probabilities over the analysis and crosses the 10% standard in only one period, $1965-69$, which is a surprising departure from the results of other papers regarding variable selection in growth models. Of the other variables that depart from historical findings the *Fertility Rate*, is perhaps the next most surprising result emerging from this analysis. Again, like the *East-Asian Dummy* Variable, this was thought to be inherently correlated with growth, being cited as one of Sala-i-Martin's (1997) core regressors, and charting high inclusion probabilities of 0.946 and 0.887 in FLS[6] and DMS[5] respectively.

While the conclusion can be drawn that certain variables are important in growth regressions, the analysis conducted in this study indicates that some variables which have been historically found to correspond to growth may not be added to the best "approximating" growth regression model with the same degree of confidence as was previously thought when the time dimension is added into the analysis. Adding the time dimension also uncovers some sluggishness in the effect of initial *Real GDP per Capita* on growth before seeing the convergence effect documented in the literature.

References

[1] Barro, R.J. (1991), Economic Growth in a Cross Section of Countries, *Quarterly Journal of Economics, 106, 407-444.*

[2] Barro, R. J. and Lee, J.-W., (1994), Data Set for a Panel of 138 Countries, *mimeo., Harvard University, Cambridge, Massachusetts*

[3] Barro, R.J., and X.X. Sala-i-Martin (1992), Convergence, *Journal of Political Economy 100, 223-251.*

[4] Fernandez, C., E. Ley, and M. F. J. Steel (2001), Benchmark Priors for *Bayesian* Model Averaging,*Journal of Econometrics, 100, 381-427.*

[5] Doppelhofer, G., Miller, R. and Sala-i-Martin, X. (2004), Determinants of long-term growth: A *Bayesian* averaging of classical estimates (BACE) approach, *American Economic Review, Vol. 94 , pp. 813-835.*

[6] Fernandez, C., E. Ley, and M. F. J. Steel (2001), Model uncertainty in cross-country growth regressions, *Journal of Applied Econometrics, vol. 16(5), pp. 563-576.*

[7] Kohn, R,, Smith, M. and Chan , D. (2001), Nonparametric Regression Using Linear Combinations of Basis Functions, *Statistics and Computing, Vol. 11, pp. 313-322.*

[8] Islam, N. (1995), Growth Empirics: A Panel Data Approach, *Quarterly Journal of Economics, Vol. 110, N0. 4, pp. 1127-1170.*

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

[10] Leamer, E. (1983), Let's Take the Con out of Econometrics. *American Economic Review, Vol. 73, 1. pp. 31-43.*

[11] Levine, R. and Renet, D. (1992) A Sensitivity Analysis of Cross-Country Growth Regressions, *American Economic Review, Vo. 82, 942-963.*

[12] Ouysse, R. and Kohn, R. (2007),*Bayesian* Variable Selection of Risk Factors in the APT Model, *Discussion Paper 2007/32, School of Economis, UNSW.*

[13] Raftery, A.E., Madigan, D. and Hoeting, J. A., (1997), *Bayesian* Model Averaging for Linear Regression, *Journal of the American Statistical Association, Vol. 92, pp. 179-191*

[14] Madigan, D., and York, J. (1995), *Bayesian* Graphical Models for Discrete Data, *International Statistical Review, Vol. 6, pp. 215-232.*

[15] Miller and Sala-i-Martin (2000)

[16] Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *Annals of Statistics, Vol. 22, pp. 1701-1762.*

[17] Sala-i-Martin, X. X. (1997), I just Ran Four Million Regressions, *Mimeo, Columbia University.*

[18] Sala-i-Martin, X. X. (1997) I just Ran Two Million Regressions, *American Economic Review, Vol. 87, pp. 178-183.*

[19] Solow, R. (1956), A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics, Vol. 70, pp. 65-94.*