Determinants of Economic Growth: A Bayesian Panel Data Approach*

Enrique Moral-Benito†

CEMFI

October 2010

Abstract

Model uncertainty hampers consensus on the key determinants of economic growth. Some recent cross-country cross-sectional analyses have employed Bayesian Model Averaging to tackle the issue of model uncertainty. This paper extends that approach to panel data models with country-specific fixed effects in order to simultaneously address model uncertainty and endogeneity issues. The empirical findings suggest that in a panel setting the most robust growth determinants are the price of investment goods, distance to major world cities, and political rights.

JEL Classification: C11, C23, O4.

Keywords: Growth determinants, model uncertainty,
Bayesian Model Averaging, dynamic panel estimation.

*I would like to thank Manuel Arellano for his overall guidance and insightful comments. I also thank Roberto León, Eduardo Ley, Luis Servén, Ignacio Sueiro and an anonymous referee for helpful comments. All errors are my own.

†E-mail address: enrique.moral@gmail.com

1 Introduction

Over the last two decades, hundreds of empirical studies have attempted to identify the determinants of growth. This is not to say that growth theories are of no use for that purpose. Rather, the problem is that different growth theories are typically compatible with one another. For example, a theoretical view holding that trade openness matters for economic growth is not logically inconsistent with another theoretical view that emphasizes the role of geography in growth. From an empirical point of view, the problem this literature faces is known as model uncertainty, which emerges because theory does not provide enough guidance to select the proper empirical model. In the empirical growth literature, the main area of effort has been the selection of appropriate variables to include in linear growth regressions, resulting in a total of more than 140 variables proposed as growth determinants.

Many researchers consider that the most promising approach to accounting for model uncertainty is to employ model averaging techniques. This approach allows constructing parameter estimates that formally address the dependence of model-specific estimates on a given model. In this context and using methods advanced by Raftery (1995), Sala-i-Martin et al. (2004) -henceforth SDM- employ the so-called Bayesian Averaging of Classical Estimates (hereafter, BACE) to determine which growth regressors should be included in linear cross-country growth regressions. In a pure Bayesian spirit, Fernandez et al. (2001a) -henceforth FLS- apply the fully Bayesian Model Averaging (BMA) approach with the same objective. This literature on BMA and growth empirics is so far based on cross-sectional studies.

The main objective of this paper is to extend the Bayesian Model Averaging methodology to a panel data framework. The use of panel data in empirical growth regressions has many advantages with respect to typical cross-country regressions. First of all, the prospects for reliable generalizations in cross-country growth regressions are often constrained by the limited number of countries available, therefore, the use of within-country variation to multiply the number of observations is a natural response to this constraint. On the other hand, the use of panel data methods allows addressing the inconsistency of empirical estimates which typically arises with omitted country-specific effects correlated with other regressors, or with endogenous variables which may be incorrectly assumed to be exogenous. Many studies such as Islam (1995) or Caselli et al. (1996) have employed panel data models with country-specific effects in empirical growth regressions.

In order to simultaneously address both omitted variable bias and issues of endogeneity, we employ a novel maximum likelihood estimator in the growth context which is able to use the within variation across time and also the between variation across countries.[1] More concretely, our likelihood function not only includes individual effects correlated with the time varying regressors but also takes into account the endogenous nature of the lagged dependent variable in our dynamic panel setting. On the other hand, we will also be able to consider two types of time-invariant country-specific heterogeneity, observable and unobservable, under the assumption that they are uncorrelated. More importantly, given the likelihood-based nature of the estimator, it can be easily combined with BMA techniques in order to also address model uncertainty.

Against this background, this paper follows Raftery (1995) and constructs weighted averages of maximum likelihood estimates. We label the approach as Bayesian Averaging of Maximum Likelihood Estimates (BAMLE), which is easy to interpret and easy to apply since, as in the version introduced by Sala-i-Martin et al. (2004), it requires only the elicitation of priors on the model space, for example through one single hyper-parameter, the expected model size, m. Moreover, the impact of different prior assumptions on the model space is minimal with the prior structure employed in this paper. This methodology is similar to the BACE approach by SDM in the sense that both follow Raftery (1995) using the Schwarz asymptotic approximation to the marginal likelihood.

The empirical findings suggest that country-specific effects correlated with other regressors play an important role since the list of robust growth determinants is not the same when we do not take into account their presence. On the other hand, once we simultaneously address model uncertainty and endogeneity issues, the empirical results indicate that the most robust growth determinants are the price of investment goods, distance to major world cities, and political rights. Finally, we also find that the fewer the candidate regressors considered the smaller the sensitivity of the empirical results to different sources of GDP data.[2] For the purpose of robustness, this suggests that the set of candidate variables should avoid inclusion of multiple proxies for the same theoretical effect.

The remainder of the paper is organized as follows. Section 2 describes the BMA methodology and explains how to extend to the panel data case the approaches employed by SDM and FLS in cross-sections. Section 3 presents the so-called BAMLE approach in order to simultaneously address model uncertainty and endogeneity issues. Firstly, it constructs the likelihood function that considers the endogeneity of the lagged dependent variable in dynamic panels. It then describes the use of the BIC approximation in the BMA context, and finally it introduces the employed prior assumptions. In Section 4 we briefly describe the data set. The empirical results are presented in Section 5. The final section concludes.

2 Bayesian Model Averaging

A generic representation of the canonical growth regression is:

$$\gamma = \theta X + \varepsilon, \tag{1}$$

where γ is the vector of growth rates, and X represents a set of growth determinants, including those originally suggested by Solow as well as others.[3] There exist potentially very many empirical growth models, each given by a different combination of explanatory variables, and each with some probability of being the 'true' model. This is the starting point of the Bayesian Model Averaging methodology.

However, there is one variable for which theory offers strong guidance, and is therefore exempt from the problem of model uncertainty: initial GDP, which should always be included in growth regressions (see Durlauf et al. (2005)). As a result, in the remainder of the paper initial GDP will be included in all models under consideration.

Using the Bayesian terminology, a model is formally defined by a likelihood function and a prior density. Suppose we have K possible explanatory variables. We will have 2^K possible combinations of regressors, that is to say, 2^K different models - indexed by M_j for $j = 1, ..., 2^K$- which all seek to explain D -the data. M_j depends upon parameters θ^j. In cases where many models are being entertained, it is important to be explicit about which model is under consideration. Hence, the posterior for the parameters calculated using M_j is written as:

$$g\left(\theta^j|D, M_j\right) = \frac{f\left(D|\theta^j, M_j\right) g\left(\theta^j|M_j\right)}{f\left(D|M_j\right)}, \tag{2}$$

and the notation makes clear that we now have a posterior, a likelihood, and a prior for each model. The logic of Bayesian inference suggests that we use Bayes' rule to derive a probability statement about what we do not know (*i.e.* whether a model is correct or not) conditional on what we do know (*i.e.* the data). This means the posterior model probability can be used to assess the degree of support for M_j. Given the prior model probability $P(M_j)$ we can calculate the posterior model probability using Bayes Rule as:

$$P(M_j|D) = \frac{f(D|M_j)\, P(M_j)}{f(D)}. \tag{3}$$

Since $P(M_j)$ does not involve the data, it measures how likely we believe M_j to be the correct model before seeing the data. $f(D|M_j)$ is often called the marginal (or integrated) likelihood, and is calculated using (2) and a few simple manipulations. In particular, if we integrate both sides of (2) with respect to θ^j, use the fact that $\int g(\theta^j|D, M_j)\, d\theta^j = 1$ (since probability density functions integrate to one), and rearrange, we obtain:

$$f(D|M_j) = \int f\left(D|\theta^j, M_j\right) g\left(\theta^j|M_j\right) d\theta^j. \tag{4}$$

The quantity $f(D|M_j)$ given by equation (4) is the marginal probability of the data, because it is obtained by integrating the joint density of (D, θ^j) given D over θ^j. The ratio of integrated likelihoods of two different models is the Bayes Factor and it is closely related to the likelihood ratio statistic, in which the parameters θ^j are eliminated by maximization rather than by integration.

Moreover, considering θ a function of θ^j for each $j = 1, ..., 2^K$, we can also calculate the posterior density of the parameters for all the models under consideration:

$$g(\theta|D) = \sum\nolimits_{j=1}^{2^K} P(M_j|D)\, g(\theta|D, M_j) \tag{5}$$

If one is interested in point estimates of the parameters, one common procedure is to take expectations across (5):

$$E(\theta|D) = \sum_{j=1}^{2^K} P(M_j|D)\, E(\theta|D, M_j). \tag{6}$$

Following Leamer (1978), we calculate the posterior variance as:

$$\begin{aligned} V(\theta|D) &= \sum_{j=1}^{2^K} P(M_j|D)\, V(\theta|D, M_j) + \\ &+ \sum_{j=1}^{2^K} P(M_j|D)\left(E(\theta|D, M_j) - E(\theta|D)\right)^2. \end{aligned} \tag{7}$$

The posterior variance in (7) incorporates not only the weighted average of the estimated variances of the individual models but also the weighted variance in estimates of the θ's across different models. This means that even if we have highly precise estimates in all the models, we might end up with considerable uncertainty about the parameter if those estimates are very different across specifications.

In words, the logic of Bayesian inference implies that one should obtain results for every model under consideration and average them using appropriate weights. However, implementing Bayesian Model Averaging can be difficult since the number of models under consideration (2^K), is often huge. This has led to various algorithms which do not require dealing with every possible model. In particular we will employ the so called Markov Chain Monte Carlo Model Composition (MC^3) algorithm (see the Computational Appendix for more details).

Given the above, we are now ready to introduce our measure of robustness. We estimate the posterior probability that a particular variable h is included in the regression, and we interpret it as the probability that the variable belongs to the true growth model. In other words, variables with high posterior probabilities of being included are considered as *robust* determinants of economic growth. This is called the *posterior inclusion probability* for variable h, and it is calculated as the sum of the posterior model probabilities for all of the models including that variable:

$$\text{posterior inclusion probability} = P\left(\theta_h \neq 0|D\right) = \sum\nolimits_{\theta_h \neq 0} P\left(M_j|D\right). \quad (8)$$

2.1 BMA and Growth Regressions

The BMA literature in the growth context (for instance, Fernandez et al. (2001a) and Sala-i-Martin et al. (2004)) is so far based on cross-sectional studies[4] in which the regressors are assumed to be strictly exogenous. Moreover, given the lack of the time series dimension in their data, those studies do not consider the existence of unobserved heterogeneity across countries. As pointed out in the introduction, it is also true that given the limited number of countries in the world the need for BMA in cross-sections is larger than in panels. This is so because in large models, cross-section regressions with 100 observations or less are not very informative and BMA provides a systematic solution to this problem. However, BMA is also relevant in panels because it allows considering the two levels of uncertainty existing in growth regressions (*i.e.* the uncertainty associated with the parameters conditional on a given model and the uncertainty in the specification of the empirical model). Therefore, the proper uncertainty measures required for inference purposes are only provided by BMA.

In this paper we extend the use of the BMA methodology to panel data models in the growth context. In subsections 2.2 and 2.3 we first consider dynamic panel data models with country-specific effects in which all the regressors and the lagged dependent variable are assumed to be strictly exogenous.[5] As a consequence, the only difference with respect to previous BMA cross-sectional studies is the presence of country-specific fixed effects correlated with the regressors.

Later in Section 3 we derive the likelihood function of dynamic panel data models with unobserved heterogeneity that relax the strict exogeneity assumption of the lagged dependent variable by using not only within variation across time but also between variation across countries. This likelihood function allows eliminating the bias associated with the within group (henceforth, WG) estimator in dynamic panels. More concretely, we adopt a correlated random effects approach in which the country-specific effects are assumed to be linearly dependent on the means (over time) of the time-varying regressors and independent of the time-invariant covariates. As we will see in Section 3, the assumptions under which this approach can accommodate unobserved heterogeneity are not more restrictive than previous approaches to this issue in the empirical growth literature. Finally, given the likelihood-based nature of the estimator, it can be easily combined with BMA techniques in order to also consider model uncertainty.

In spite of the focus on robustness of the BMA approach, Ley and Steel (2009) show that the results are fairly sensitive to the use of different prior assumptions. In this paper we employ the hierarchical priors over the model size[6] proposed by Ley and Steel (2009) in order to minimize the effect of weakly-held prior views.

On the other hand, Ciccone and Jarocinski (2009) conclude that the list of growth determinants emerging from BMA approaches is sensitive to arguably small variations in the international GDP data used in the estimations. In an attempt to investigate this issue, we replicate our exercises with four different sources of GDP data: the three last versions of the Penn World Table (*i.e.* PWT 6.1, PWT 6.2 and PWT 6.3) and the GDP data reported in the World Development Indicators from the World Bank. We also consider different numbers of candidate regressors in our replications.

2.2 BACE Approach in a Panel Data Context

The combination of the WG estimator with BMA techniques is the simplest and most natural extension to panel data models of previous BMA approaches in the growth context. Therefore in this subsection we show how to apply the WG estimator in the BMA framework in the spirit of Raftery (1995). In particular, the only difference with respect to the BACE approach by Sala-i-Martin et al. (2004) is the inclusion of country-specific effects (*i.e.* unobserved heterogeneity). However, it is important to remark that given the well-known WG bias in short dynamic panels, we will subsequently adopt an alternative approach that addresses this issue (see Section 3).

In the panel data context, for a given group of regressors, that is, for a given model M_j, the estimated econometric model consists of the following equation and assumptions:

$$y_{it} = \alpha y_{it-\tau} + x_{it}'^{j}\beta^{j} + \eta_i + \zeta_t + v_{it} \quad (t = 1, ..., T) \; (i = 1, ..., N) \tag{9}$$

$$E\left(v_i | y_i, x_i^j, \eta_i\right) = 0, \tag{A1}$$

$$Var\left(v_i | y_i, x_i^j, \eta_i\right) = \sigma^2 I_T. \tag{A2}$$

where $v_i = (v_{i1}, ..., v_{iT})'$, $x_i^j = \left(x_{i1}^j, ..., x_{iT}^j\right)'$ and $y_i = (y_{i1}, ..., y_{iT})'$. We observe y_{it} (the log of per capita GDP for country i in period t) and the k^jx1 vector of explanatory variables x_{it}^j included in model M_j, but not η_i, which is the time-invariant component of the error term capturing the unobserved heterogeneity.

Although under assumptions (A1) and (A2) the WG estimator is the optimal estimator of α and β^j, it is now well known that in dynamic panels with small T (as will be the case in this paper) the WG is badly biased because assumption (A1) does not hold by definition (see Nickell (1981)). In the next section we will relax assumptions (A1) and (A2) in order to address this issue.

Note that in addition to the individual specific fixed effect η_i, we have also included the term ζ_t in (9). That is to say, we are including time dummies in the model in order to capture unobserved common factors across countries and therefore we are not ruling out cross-sectional dependence. In practice, this is done by simply working with cross-sectional de-meaned data. In the remaining of the exposition, we assume that all the variables are in deviations from their cross-sectional mean.

Following Raftery (1995) and Sala-i-Martin et al. (2004) we have implemented the so-called BACE approach in this context. The idea of BACE is to assume diffuse priors (as an indication of our ignorance) and make use of the result that, in the linear regression model, for a given model M_j, standard diffuse priors and Bayesian regression yield posterior distributions identical to the classical sampling distribution of OLS.

Under the assumptions stated above we can rewrite (6) as:

$$E\left(\theta|y\right) = \sum\nolimits_{j=1}^{2^K} P\left(M_j|y\right) \widehat{\theta}^j, \tag{10}$$

where $\widehat{\theta}^j$ is the WG estimate for θ with the regressor set that defines model j, and y refers to the data. Moreover, as the posterior odds' behavior is problematic with diffuse priors, Raftery (1995) proposes instead the use of the Schwarz asymptotic approximation to the integrated likelihood,[7] and therefore:

$$P\left(M_j|y\right) = \frac{P\left(M_j\right)\left(NT\right)^{-k^j/2} SSE_j^{-(NT)/2}}{\sum_{i=1}^{2^K} P\left(M_i\right)\left(NT\right)^{-k^i/2} SSE_i^{-(NT)/2}}, \tag{11}$$

where NT is the number of observations, K is the total number of regressors, k^j is the number of parameters included in model j and SSE_j is the sum of squared residuals of the j-model's regression.

In the case of balanced panels the number of observations in (11) is given by NT because the WG log-likelihood function can be written as a sum of NT contributions (see for example Arellano (2003)). Therefore the curvature of the log-likelihood function grows at the rate NT, and this growth rate is the quantity that should appear in the penalty term in (11) as suggested by Kass and Wasserman (1995). For unbalanced panels, as long as all the models are estimated with the same observations regardless of the variables included, one possibility is to use the number of observations employed in the estimation (*i.e.* $\sum_{i=1}^{N} T_i$ where T_i is the number of time series observations for individual i).

On the other hand, we do not include the number of fixed effects in k^j (the number of parameters in model j) since the log-likelihood version of the WG estimator can be written as a function of only α, β and σ^2. In any event, all the considered models allow for N fixed effects, and thus the inclusion or not of N in the number of parameters would not have any effect on either the posterior model probabilities or the final results.

Regarding the priors on the model size (W), the BACE approach assumes that each variable is independently included in a model:

$$W \sim Bin\left(K, \xi\right) \tag{12a}$$

$$E\left(W\right) = K\xi \Rightarrow \xi = \frac{m}{K}. \tag{12b}$$

Note that with this prior structure, the researcher only needs to fix the prior expected model size $E(W) = m$ which determines the prior inclusion probability (ξ) through (12b). On the other hand, the researcher can also fix the prior inclusion probability that will imply the prior expected model size, as we will see in the next subsection. In particular, Sala-i-Martin et al. (2004) propose $m = 7$ as a reasonable prior mean model size in the cross-country context. Here, we propose $m = 5$ for the panel data case because previous studies on panel growth regressions typically consider fewer covariates than cross-sectional studies because of the lack of time series information for some variables. In any case, as we will see, different prior assumptions about the value of m have practically no effect on the results with the prior structure we will employ, so the choice of m is not critical for the data set used in this paper.

2.3 BMA-FLS Approach in a Panel Data context

One question arises when we think in terms of Bayesian econometrics: how sensitive are the results to the choice of priors by the researcher? In this section, instead of the BACE approach based on diffuse priors, we briefly review the full Bayesian approach with the benchmark priors proposed by Fernandez et al. (2001b). These priors can be easily applied to the panel data case (fixed-effects model) if we rewrite the M_j model in the previous section as:

$$y_{it} = \alpha y_{it-\tau} + x_{it}^{\prime j}\beta^{j} + \phi_1 D_1 + ... + \phi_N D_N + \zeta_t + v_{it} \; (t = 1, ..., T) \; (i = 1, ..., N), \quad (13)$$

where the coefficients $(\phi_1 ... \phi_N)$ are the individual unobservable effects for each country, $(D_1 ... D_N)$ are N dummy regressors and again, all variables will be in deviations from their cross-sectional means given the presence of the time dummy ζ_t. Assumptions (A1) and (A2) are assumed (incorrectly) to hold here, and the error term is supposed to follow a normal distribution. Fernandez et al. (2001b) propose a natural conjugate prior distribution which allows employment of the exact Bayes factor instead of using asymptotic approximations. For the variance parameter, which is common for all the models under consideration, the prior is improper and non-informative:

$$p(\sigma) \propto \sigma^{-1}. \quad (14)$$

The g-prior (Zellner (1986)) for the slope parameters is a normal density with zero mean and covariance matrix equal to:

$$\sigma^2 \left(g_0 Z^{\prime j} Z^{j} \right)^{-1}, \quad (15)$$

where $Z^j = (y_{-1}, x^j, D_1, ..., D_N)$ and:

$$g_0 = \min\left(\frac{1}{NT}, \frac{1}{(k^j)^2}\right).$$

With this prior, both the posterior for each model and the Bayes factor have a closed form. Concretely, the Bayes factor (the ratio of integrated likelihoods) for model M_j versus model M_i is given by:

$$B_{ji} = \left(\frac{g_{oj}}{1+g_{oj}}\right)^{\frac{k_j+1}{2}} \left(\frac{g_{oi}+1}{g_{oi}}\right)^{\frac{k_i+1}{2}} \left(\frac{\frac{1}{g_{oi}+1}SSE_i + \frac{g_{oi}}{g_{oi}+1}(y'y)}{\frac{1}{g_{oj}+1}SSE_j + \frac{g_{oj}}{g_{oj}+1}(y'y)}\right)^{\frac{NT}{2}}. \tag{16}$$

Once we have specified the distribution of the observables given the parameters and the prior for these parameters, we only need to define the prior probabilities for each of the models. In particular, FLS assume that every model has the same *a priori* probability of being the true model:

$$P(M_j) = 2^{-K}. \tag{17}$$

The prior in (17) was also considered in Raftery (1995) and it is the Binomial prior previously described but implicitly employing $m = K/2$ instead of $m = 7$. Therefore both priors on the model space can be interpreted in terms of the Binomial prior that only requires the elicitation of one hyper-parameter.

2.4 On the Effect of Prior Assumptions

We have presented and described two different prior structures employed in the BMA context. Both approaches give very similar results, and this is often misinterpreted as a symptom of robustness with respect to prior assumptions. Ley and Steel (2009) show that this similarity arises mostly by accident. The reason is that the different choices of the prior inclusion probability of each variable (ξ) – treated as fixed in both approaches – compensate the different penalties to larger models implied by the diffuse priors of SDM and the informative g-priors of FLS.

The effect of weakly-held prior views (such as those that apply in the growth regression context) should be minimal. In search of this minimal effect, Ley and Steel (2009) propose a hierarchical prior over model size (W) given by:

$$W \sim Bin\,(K, \xi) \tag{18}$$

$$\xi \sim Be\,(a, b)\,, \tag{19}$$

where $a, b > 0$ are hyper-parameters to be fixed by the researcher. The difference with respect to SDM and FLS is to make ξ random rather than fixed. Model size W will then satisfy:

$$E\,(W) = \frac{a}{a+b}K. \tag{20}$$

The model size distribution generated in this way is the so-called Binomial-Beta distribution. Ley and Steel (2009) propose to fix $a = 1$ and $b = (K - m)/m$ through equation (20), so we only need to specify m, the prior mean model size, as in the previous approaches.

As shown by Ley and Steel (2009), this prior specification with ξ random rather than fixed implies a substantial increase in prior uncertainty about model size, and makes the choice of prior model probabilities (for instance through m) much less critical. Moreover, as we will later see in Table 1, with random ξ the effects of different prior assumptions are much less pronounced.

3 Bayesian Averaging of Maximum Likelihood Estimates (BAMLE)

So far we have applied model averaging techniques to panel growth regressions with country-specific effects but assuming strict exogeneity of all the right hand side variables (*i.e.* we have not addressed the endogeneity of the lagged dependent variable in dynamic panels). We will now construct a likelihood function that allows us to address this issue. Then we will combine the resulting maximum likelihood estimator with BMA techniques using the BIC approximation in the so-called BAMLE approach.

Following Raftery (1995), the BAMLE approach is based on averaging maximum likelihood estimates in a Bayesian spirit, *i.e.*, we rewrite equation (6) as follows:

$$E\left(\theta|y\right)=\sum\nolimits_{j=1}^{2^K}P\left(M_j|y\right)\widehat{\theta}_{ML}^{j}. \tag{21}$$

where $\widehat{\theta}_{ML}^{j}$ is the maximum likelihood estimate for θ in model j.[8]

The argument behind equation (21) is twofold: *(i)* assuming diffuse priors on the parameter space of a given model, the posterior mode coincides with the MLE. *(ii)* in large samples, for any given prior, the posterior mode is very close to the MLE and then equation (21) would only hold as an approximation.

Therefore, if we face a situation with either no prior information and any sample size or any informative prior and a large sample, we can avoid the need to specify priors over model parameters in ways that might prove controversial by using a maximum likelihood estimator.

3.1 The Likelihood Function

The panel data methods employed in Section 2 only permit the use of the within variation in the data. This causes two main drawbacks: (i) Since Nickell (1981) it is well-known that given assumption (A1) does not hold in dynamic panels, the WG estimator of α is biased when T is small, as will be our case. Given the importance of this parameter -the convergence parameter- in the growth context, it is desirable to get a fixed T, large N consistent estimator of α. (ii) Given the required within groups transformation, we cannot exploit the information contained in regressors without time variation. This situation implies that we are not considering all the potential determinants of economic growth. For instance, some theories argue that geographic factors without time variation matter for growth.

Given the Bayesian spirit of the approach, we propose here to use a maximum likelihood estimator - for a given model - which permits solving the two problems just described.

For a given model M_j we can write:

$$y_{it} = \alpha y_{it-\tau} + x_{it}'^{j}\beta^{j} + z_i^{j}\gamma^{j} + \eta_i + \zeta_t + v_{it} \tag{22}$$

Moreover, we can go further and assume:[9]

$$v_{it}|y_{it-1}...y_{i0}, x_i^j, z_i^j, \eta_i \sim N\left(0, \sigma_v^2\right) \tag{A3}$$

$$\eta_i|y_{i0}, x_i^j, z_i^j \sim N\left(\varphi y_{i0} + \delta^j \overline{x}_i^j, \sigma_\eta^2\right) \tag{A4}$$

where $\overline{x}_i^j$ is the time-series mean[10] of x^j for individual i ($\overline{x}_i^j = (1/T)\sum_{t=1}^{T} x_{it}^j$). Note that in (A3) we are relaxing the assumption of strict exogeneity of the lagged dependent variable (*i.e.* we allow that current shocks affect future values of the dependent variable as implied by the dynamics of the model). This is the key assumption to obtain fixed T, large N consistent estimates of the autoregressive parameter α in (22).

Under assumptions (A3) and (A4) we can write the likelihood as:[11]

$$\begin{aligned}\log f\left(y_i|y_{i0}, x_i^j, z_i^j\right) \propto & -\frac{T-1}{2}\log\sigma_v^2 - \\ & - \frac{1}{2\sigma_v^2}\left(y_i^* - \alpha y_{i(-1)}^* - x_i^{*j}\beta^j\right)'\left(y_i^* - \alpha y_{i(-1)}^* - x_i^{*j}\beta^j\right) - \\ & - \frac{1}{2}\log\omega^2 - \frac{1}{2\omega^2}\left(\overline{y}_i - \alpha\overline{y}_{i(-1)} - \gamma^j z_i^j - \phi^j\overline{x}_i^j - \varphi y_{i0}\right)^2,\end{aligned} \quad (23)$$

where $\phi^j = \beta^j + \delta^j$, φ and ω^2 are the linear projection coefficients of $\overline{u}_i$ on $\overline{x}_i^j$ and y_{i0}. $\overline{u}_i$ is equal to $\eta_i + \overline{v}_i$, and $\overline{v}_i = (1/T)\sum_{t=1}^T v_{it}$. Moreover, y_i^*, $y_{i(-1)}^*$ and x_i^{*j} denote orthogonal deviations of y_i, $y_{i(-1)}$ and x_i^j respectively.

Thus, the Gaussian log-likelihood given y_{i0}, x_i^j and z_i^j can be decomposed into a within-group and a between-group component. This allows us to obtain a fixed T, large N consistent estimator for α (Alvarez and Arellano (2003)). Furthermore, the between-group component together with the orthogonality assumption between z_i^j and η_i allow for identification of γ^j.

It is important to remark here that the resulting maximum likelihood estimator is consistent and asymptotically normal regardless of non-normality. More specifically, our first order conditions correspond to a Generalized Method of Moments (GMM) problem with a convenient choice of weighting matrix (see Arellano (2003) pp.71-73). Therefore our approach to unobserved heterogeneity is as robust as panel GMM estimators under time-series homoskedasticity.[12]

We should emphasize that assumption (A4) implies that the regressors with and without temporal variation are treated differently. While the x's can be correlated with the unobservable fixed effect, the z's are independent. One interpretation is that, in addition to the traditional unobserved heterogeneity between countries given by the η_i term, there also exists a second type of fixed but observable heterogeneity given by the z_i variables. Moreover, both types of heterogeneity must be uncorrelated. For instance, we may think about observable geographic factors such as land area, which are assumed to be independent from unobservables of each country such as the ability of its population. With the BAMLE approach, we will be able to conclude which observable fixed factors are more important in promoting economic growth. This conclusion could also be obtained by using standard random effects estimation, but it is important to remark that with our approach we do not need to assume independence between the country specific effect and time varying regressors, which seems to be implausible in this context.

3.2 The BIC Approximation

Once we have specified the likelihood function of the data, we need a few more ingredients for the implementation of the BAMLE methodology. An essential one is the derivation of the integrated likelihood for a given model presented in equation (4). Various analytic and numerical approximations have been proposed to address this problem. Following Raftery (1995) and Sala-i-Martin et al. (2004) we will make use of the Bayesian Information Criterion (BIC) approximation, which is both simple and widely used.

We can approximate the Bayes factor between models M_i and M_j, $B_{ij} = \frac{f(y|M_i)}{f(y|Mj)}$ such that (Raftery (1995)):

$$S = \log f\left(y|\widehat{\theta}_i, M_i\right) - \log f\left(y|\widehat{\theta}_j, M_j\right) - \frac{(k_i - k_j)}{2}\log\left(NT\right), \tag{24}$$

where $\widehat{\theta}_i$ is the MLE under M_i, k_i is the dimension of $\widehat{\theta}_i$ (which does not include the N effects in the case of the likelihood function proposed in the previous subsection), and NT is the sample size for balanced panels (see Section 2.2 for a more detailed discussion). As $NT \to \infty$, this quantity, often called the Schwarz criterion, satisfies:

$$\frac{S - \log B_{ij}}{\log B_{ij}} \to 0 \tag{25}$$

Minus twice the Schwarz criterion is often known as the Bayesian information criterion (BIC):

$$BIC = -2S = -2\log B_{ij}. \tag{26}$$

Although the relative error of $exp(S)$ in approximating B_{ij} is generally $O(1)$, Kass and Wasserman (1995) show that under a reasonable choice of priors[13] the error is $O(n^{-1/2})$ instead of $O(1)$. This error is much smaller and tends to zero as the sample size increases.

The value of BIC for model M_j denoted BIC_j, is the approximation to $2logB_{0j}$ given by (26), where B_{0j} is the Bayes factor for model M_0 against M_j, where M_0 could be the null model with no independent variables. Moreover, we can manipulate the previous equations in the following manner:

$$\begin{aligned} B_{ij} &= \frac{f(y|M_i)}{f(y|M_j)} = \frac{\frac{f(y|M_i)}{f(y|M_0)}}{\frac{f(y|M_j)}{f(y|M_0)}} = \frac{B_{i0}}{B_{j0}} = \frac{B_{0j}}{B_{0i}}. \\ 2\log B_{ij} &= 2\left[\log B_{0j} - \log B_{0i}\right] = BIC_j - BIC_i. \end{aligned}$$

In addition, we can rewrite equation (3) as:

$$\begin{aligned} P(M_j|y) &= \frac{f(y|M_j)P(M_j)}{\sum_{i=1}^{2^K} f(y|M_i)P(M_i)} = \\ &= \frac{\frac{f(y|M_j)}{f(y|M_h)} f(y|M_h)P(M_j)}{\sum_{i=1}^{2^K} \frac{f(y|M_i)}{f(y|M_h)} f(y|M_h)P(M_i)} = \\ &= \frac{B_{jh} f(y|M_h)P(M_j)}{\sum_{i=1}^{2^K} B_{ih} f(y|M_h)P(M_i)} = \\ &= \frac{B_{j0}P(M_j)}{\sum_{i=1}^{2^K} B_{i0}P(M_i)}, \end{aligned} \tag{27}$$

where the last equality holds because $B_{00} = 1$ and $BIC_0 = 0$. Moreover, since $BIC_j = 2\log B_{0j} = 2\log(1/B_{j0})$ then $B_{j0} = \exp(-0.5 * BIC_j)$.

Given the above, instead of integrating to obtain the marginal likelihood in

(4), we will use the following result:

$$f\left(y|M_j\right) \propto \exp\left(-\frac{1}{2}BIC_j\right), \tag{28}$$

and therefore:

$$P\left(M_j|y\right) = \frac{P\left(M_j\right)\exp\left(-\frac{1}{2}BIC_j\right)}{\sum_{i=1}^{2^K} P\left(M_i\right)\exp\left(-\frac{1}{2}BIC_i\right)}. \tag{29}$$

Furthermore, the posterior odds (*prior odds* x *Bayes Factor*) becomes:

$$\frac{P\left(M_i|y\right)}{P\left(M_j|y\right)} = \frac{P\left(M_i\right)}{P\left(M_j\right)}\frac{\exp\left(-\frac{1}{2}BIC_i\right)}{\exp\left(-\frac{1}{2}BIC_j\right)}. \tag{30}$$

3.3 The Choice of Priors

Bayesian inference may be controversial because it requires specification of prior distributions which are subjectively chosen by the researcher. Moreover, Bayesian calculations may be extremely hard and computationally demanding when estimating millions of non-regular models.

Given the use of a maximum likelihood estimator and the BIC approximation, BAMLE avoids the need to specify a particular prior for the parameters of a given model.

As a result, for the implementation of BAMLE, the researcher only needs to specify priors on the model space. In particular, in an attempt to limit the effects of weakly held prior views, we suggest to employ the Binomial-Beta prior structure proposed by Ley and Steel (2009), as described in the previous section.

4 Data

A huge number of variables have been proposed as growth determinants in the cross-country literature, including variables with and without time variation. However, data for many of the latter are not available over the entire sample period under consideration in this paper.[14] Since our main goal is to work with a panel data set, we limit our selection of time-varying variables to those for which data are available over the entire period 1960-2000.

In the construction of our data set, we have considered two different criteria. The first selection criterion derives from our aim of obtaining comparable results with the existing literature, and the second criterion comes from the fact that we need to work with a balanced panel.

With these restrictions, the total size of our data set becomes 35 variables (including the dependent variable, the growth rate of per capita GDP) for 73 countries and for the period 1960-2000. In order to lessen the problem of serial correlation in the transitory component of the disturbance term, we have split our sample in five year periods. Therefore we have eight observations for each country, that is to say, we have a sample of 584 observations.

Among the 19 regressors with temporal variation in our data set, there are both stock and flow variables. Following Caselli et al. (1996), stock variables such as population and years of primary education are measured in the first year of each five-year period. On the other hand, flow variables such as population growth and investment rate are measured as five-year averages.

4.1 Determinants of Economic Growth

The augmented Solow model can be taken as the baseline empirical growth model. It comprises four determinants of economic growth, initial GDP, rates of human and physical capital accumulation, and population growth. We capture these growth determinants through the ratio of real investment to GDP from PWT version 6.2, the stock of years of education from Barro and Lee, and demographic variables such as life expectancy from the World Bank, the ratio of labor force to total population and population growth from PWT 6.2. In addition to those four determinants, the Durlauf et al. (2005) survey of the empirical growth literature identifies 43 distinct growth theories and 145 proposed regressors as proxies; each of these theories is found to be statistically significant in at least one study. Due to data availability, our set of growth determinants is a subset of that identified by Durlauf et al. (2005). We consider the three broad variable categories below.

- Macroeconomic and external environment: Following Easterly (1993), we consider the investment price level (i.e., the PPP investment deflator from PWT 6.2) as a proxy for the level of distortions that exists in the economy. We also consider the size of the government measured by the ratio of government consumption to GDP from PWT 6.2. Many authors such as Barro (1991) have considered this ratio as an additional measure of distortions in the economy. The argument is that government consumption has no direct effect on private productivity but lowers saving and growth through the distorting effects from taxation or government-expenditure programs. On the other hand, the trade regime/external environment is captured by the degree of trade openness, measured by imports plus exports as a share of GDP from PWT 6.2. Many authors such as Levine and Renelt (1992) have considered this ratio. However, this measure is sometimes criticized because it only takes into account the volume of trade and not the nature of trade policies in a given country. In order to capture the degree of openness as a proxy for distortions in trade policies, we also consider an alternative indicator, the SW openness index constructed by Sachs and Warner (1995). It is worth mentioning that the SW indicator has its own limitations as pointed out by Rodriguez and Rodrik (2000). Considering these two measures we aim to conclude which dimension of trade openness matters most.

- Governance and institutions: The understanding of the role of democracy and institutions in the process of economic growth has generated an enormous amount of research. In this paper we examine the hypothesis that political freedom and institutional quality are significant determinants of economic growth using political rights and civil liberties indices to measure the quality of institutions and capture the occurrence of free and fair elections and decentralized political power. Both indices are constructed by Freedom House, and they are freely available at http://www.freedomhouse.org. Barro and Lee (1994) and Sala-i-Martin (1997) among others considered these indices as proxies of the quality of institutions and governance.

- Geography and fixed factors: Since the seminal paper by Sachs and Warner (1997) there is an influential view arguing that differences in natural endowments, such as climatic conditions, can account for income differences across countries. Very closely related, another view stresses market proximity (non-remoteness) in explaining spatial variation in economic activity, as emphasized in the literature on new economic geography following Krugman (1991). In order to examine the extent to which geography matters for growth, we use a variety of geographic indicators such as the percentage of land area in the geographical tropics or the fraction of population in geographical tropics. On the other hand, as proxies for remoteness we use, among others, the minimum distance to New York, Rotterdam or Tokyo, the fraction of land area near navigable water and a dummy for landlocked countries. Finally, other fixed but not geographic factors such as active participation in conflicts during the sample period (a war dummy) or the timing of independence, may have an effect on economic growth as pointed out by Barro and Lee (1994) and Gallup et al. (2001) respectively. The geographical variables and fixed factors considered in this paper were all taken from the Center for International Development (CID) at Harvard University.

A list of variables with their corresponding description and sources can be found in the Data Appendix, as well as the list of countries included in the sample.

5 Results

5.1 Panel BACE-SDM and Panel BMA-FLS Results

Table 1 reports the posterior inclusion probability of the 19 regressors with time variation included in our data set after applying BACE-SDM and BMA-FLS procedures in a panel data context. The table highlights the sensitivity of the results to the different prior assumptions. Concretely, comparison of columns 1 and 3, and 2 and 4, shows that with fixed ξ different assumptions about the prior mean model size, $m = 5$ or $m = K/2$, generate quite different posterior inclusion probabilities. More specifically, when we do not penalize larger models in any way – that is to say, when we employ $m = K/2$ instead of $m = 5$ in the BACE-SDM approach (columns 3 and 1 respectively) – the posterior inclusion probabilities are higher. On the other hand, when we do penalize bigger models in both ways employing $m = 5$ in the BMA-FLS approach (column 2), the posterior inclusion probabilities are smaller. This also highlights the "fortuitous robustness" which emerges when we compare the BMA-FLS and BACE-SDM results in columns 1 and 4, that is to say, different prior assumptions on model size have substantial effects on the results. Furthermore, analyzing columns 5 to 8 of Table 1, we can conclude that the effects of prior assumptions on model size are much less important in the case of random ξ (*i.e.* the hierarchical priors over the model size proposed by Ley and Steel (2009)). Moreover, the last row of the table indicates that expected model size should be close to 5 in the panel data framework.

Table 2 summarizes the posterior distributions of the parameters corresponding to the 19 variables of our data set with time variation when we apply the

BACE-SDM and BMA-FLS approaches with country-specific effects. In particular, it reports the posterior inclusion probability, the posterior mean and the posterior standard deviation of these distributions. These results are based on the whole sample, that is, 73 countries for the period 1960-2000. The main conclusion from the table is that, in addition to initial GDP, there are several covariates which appear to be robustly associated with economic growth. However these covariates are in general not the same as those emerging in the cross-sectional case as in Sala-i-Martin et al. (2004) and Fernandez et al. (2001a). This is an indication that country-specific effects matter and make a difference in this respect. Note however, that given the nature of our paper, the main conclusions will be obtained according to the results presented in the next subsection (Table 3).

5.2 BAMLE Results

Results when applying the BAMLE approach with GDP data from PWT 6.2 for the whole period are summarized in Table 3. Additionally to initial GDP, a fair number of regressors could be considered as robust determinants of economic growth accordingly to the Bayesian robustness check used in the approach. The most conclusive evidence is for investment price, air distance to big cities and political rights. All the three regressors affect growth with the expected sign: a low level of distortions in the economy (*i.e.* lower investment price), a better geographic situation and a higher level of democracy (*i.e.* lower value for the political rights index) would promote economic growth. This suggests that growth-promoting policy strategies should aim to reduce taxes and distortions that raise the prices of investment goods, and promote democracy-enhancing institutional reforms. On the other hand, since their posterior inclusion probabilities are higher than their prior inclusion probabilities, many other variables such as demographic indicators, a measure of trade openness, the dummy for landlocked countries, the investment share, the civil liberties index and the government consumption share can be considered as robust determinants of economic growth.

Although the comparison between posterior inclusion probabilities and prior inclusion probabilities has been commonly used in the economics BMA literature, it must be interpreted with care. Even if the posterior inclusion probability is lower than the prior inclusion probability for a given variable, it might be the case that this particular variable is important to decision-makers under some circumstances. Therefore, although useful for presentation purposes, the mechanical application of a threshold, or a simple comparison between the prior and the posterior, should often be avoided.

Finally, there is one regressor, life expectancy, that poses a puzzle. In spite of having the highest posterior inclusion probability, we think it cannot be viewed as robust because its posterior standard deviation is bigger than its posterior mean. This means that this variable is associated with economic growth, but we cannot conclude in which direction because of the model uncertainty problem.

As pointed out by Temple (1998) among others, one important concern in empirical growth regressions is the presence of outliers (*i.e.* observations measured with a substantial degree of error or drawn from a different regime). If atypical observations are present in our data, they might have an unduly large influence on the results. In order to check the presence of influential observations in our results we use the influence statistics for panel data models proposed by Banerjee and Frees (1997). Intuitively, we first estimate a model with the full sample and then, we re-estimate the model N more times by deleting one country at a time. With the $N+1$ estimates we construct the Banerjee and Frees' statistic for each country and test if any of them is influential. The computed statistics for the four variables with highest posterior inclusion probabilities (PIP) are plotted in Figure 1. Under the null of no influence these statistics are approximately distributed as a χ^2 with one degree of freedom. Since the 5% critical value is 3.84 we can conclude that there is no country in our sample exerting special influence on the results[15] (at least on the variables that we label as robust).

It is worth mentioning that the posterior mean conditional on inclusion of the lagged dependent variable (initial GDP) in Table 3 implies a rate of conditional convergence of $\lambda = 0.006$. This rate of convergence is much lower than the one found in previous panel studies such as Caselli et al. (1996).[16] Moreover, the high standard deviation suggests that previous results in the literature should be interpreted with care. More concretely, it indicates that once we control for model uncertainty and other potential inconsistencies (*e.g.* omitted variable and endogeneity biases), the data cannot precisely identify the rate of convergence.

5.3 Sensitivity Results

Ciccone and Jarocinski (2009) point out that agnostic BMA approaches lead to conclusions that are sensitive across available sources of international GDP data. They compare the different Posterior Inclusion Probabilities (PIP) emerging when using alternative sources of data on GDP. For each variable j they estimate its PIP using PWT 6.1 and PWT 6.2 GDP data. Then they compute the absolute value of the difference between the two PIP's (abs_diff_j). From their set of 67 explanatory variables they conclude that the differences are substantial.

In an attempt to further investigate this issue, Table 4 presents measures of sensitivity of the results when using different sources of international GDP data. We compare the results obtained with the baseline GDP data used in this paper (PWT 6.2) with another two versions of the Penn World Table (PWT 6.1 and PWT 6.3) and the GDP data reported in the World Development Indicators 2005 from the World Bank. More concretely, Table 4 reports the average and median of the abs_diff for all the variables.

The number of explanatory variables considered (K) seems to be a key determinant of the sensitivity. We can see that in all the comparisons, both the average and the median sensitivity are smaller when considering 10 regressors instead of 19. This result implies that the fewer the regressors considered the smaller the sensitivity.[17] For the sake of robustness, this result suggests that the set of candidate variables should avoid inclusion of multiple proxies for the same theoretical effect.

Another important result from Table 4 is that the sensitivity when comparing PWT 6.2 and PWT 6.3 in Panel B is smaller than in Panel A when we compare PWT 6.2 and PWT 6.1. Therefore, the last available revision of the Penn World Table seems to produce more stable results than previous revisions. On the other hand, results using WDI 2005 and PWT 6.2 data are more sensitive than across different versions of the PWT.

6 Concluding Remarks

In spite of a huge amount of empirical research, the drivers of economic growth are not well understood. This paper attempts to provide insights on the growth puzzle by extending the Bayesian Model Averaging (BMA) approach to a panel data setting. Based on Raftery (1995), we employ the so-called Bayesian Averaging of Maximum Likelihood Estimates (BAMLE) method in a panel data framework to determine which variables are significantly related to growth. Similarly to the BACE approach, this method does not require the specification of prior distributions for the parameters of every model under consideration, and it only involves priors on the model space (for instance through one hyper-parameter, the expected model size m). Moreover, the BAMLE approach introduces two improvements with respect to previous model-averaging and robustness-checking methods applied to empirical growth regressions: (i) it addresses the problem of inconsistent empirical estimates by using a dynamic panel estimator, and (ii) it minimizes the impact of prior assumptions about the only hyper-parameter in the approach by employing the Binomial-Beta priors on the model space proposed by Ley and Steel (2009).

The empirical findings suggest that country-specific effects correlated with other regressors play an important role since the list of robust growth determinants is not the same when we do not take into account their presence. Our results indicate that once model uncertainty and other potential inconsistencies are accounted for, there exist economic, institutional, geographic and demographic factors robustly correlated with growth. The most robust determinants are the price of investment as an indication of the level of distortions in the economy, the distance to big cities as a proxy for remoteness, and the institutional framework proxied by the political rights index. Other variables which can be considered as robust include demographic factors (population growth, urban population and population), geographical dummies (such as the dummy for landlocked countries), measures of openness and civil liberties, and macroeconomic indicators such as the investment share and the government consumption share.

On the other hand, our empirical point estimate of the rate of convergence, after controlling for both model uncertainty and endogeneity of the lagged dependent variable, is much lower than the one typically found in panel studies and surprisingly similar to that commonly found in cross-section studies. Moreover, looking at the whole posterior distribution of the convergence coefficient we observe that there is a significant amount of probability mass on both sides of zero. Therefore, one would conclude that there is no evidence of conditional convergence according to this result, or, more precisely, that the evidence is ambiguous because the data cannot accurately identify the rate of convergence.

As a final remark, it is worth mentioning that the dynamic panel estimator proposed in this paper addresses the endogeneity of regressors with time variation with respect to the permanent component of the error term as well as the endogeneity of the lagged dependent variable with respect to the transitory component of the error term. However, many other regressors such as the labor force or the investment share should ideally be considered as predetermined instead of strictly exogenous with respect to the transitory component of the error term, and this point remains unresolved in the BMA context. Hence, the estimates might change under less stringent exogeneity assumptions. This issue is left for future research.

Notes

[1]This maximum likelihood estimator can be described as a correlated random effects estimator as in the work of Mundlak (1978) and Chamberlain (1984).

[2]Ciccone and Jarocinski (2009) point out that the results emerging from agnostic model averaging approaches are sensitive to small variations in the international GDP data used (*e.g.* different versions of the Penn World Table (PWT)).

[3]The inclusion of additional control variables to the regression suggested by the Solow (or augmented Solow) model can be understood as allowing for predictable and additional heterogeneity in the steady state

[4]Chen et al. (2009) propose a pseudo BMA approach in a panel data context. In particular they compute weighted averages of GMM estimates using as weights the Schwarz asymptotic approximation but replacing the fully specified likelihood by the exponentiated GMM objective function. In a follow-up paper, Mirestean and Tsangarides (2009) apply this methodology to growth regressions.

[5]Since the lagged dependent variable in dynamic panels is not strictly exogenous by definition, later in the paper we will present how to address this issue in the BMA framework.

[6]There is a one-to-one mapping from priors over the model size to priors over the inclusion probabilities of the regressors (see Section 2.4 for more details).

[7]For a more detailed discussion of the use of this asymptotic approximation in the BMA context see Section 3.2.

[8]For its use in the BMA context $\widehat{\theta}^j_{ML}$ must be considered as a maximum likelihood estimate (MLE). However, from a frequentist viewpoint the same estimate can be interpreted as a pseudo MLE for single-model estimation purposes.

[9]Note that all data will be cross-sectional de-meaned given the inclusion of

time dummies.

[10]We consider the means over time in the spirit of Mundlak (1978) instead of the full vector of time-series observations à la Chamberlain to avoid the proliferation of coefficients.

[11]See Alvarez and Arellano (2003) for the demonstration in the pure autoregressive model. We add here additional exogenous explanatory variables with and without temporal variation.

[12]Ahn and Schmidt (1995) discuss GMM estimators of this kind.

[13]A prior on the parameter space that is a multivariate normal with mean equal to the MLE of the parameters and variance equal to the inverse of the expected Fisher information matrix for one observation. This prior is usually called the Unit Information Prior.

[14] For instance, the fraction of GDP in mining and the fraction of Muslim population (both considered in Fernandez et al. (2001a) and Sala-i-Martin et al. (2004)) are only available for the year 1960.

[15]Due to its single-case nature and as a result of masking, the employed deletion diagnostic can fail in the presence of multiple unusual countries jointly influencing the results.

[16]This results suggests that panel studies based on first differenced GMM estimators, where the estimated rate of convergence is surprisingly high, suffer from finite-sample biases. In fact, by resorting to auxiliary stationarity assumptions in the GMM framework, Bond et al. (2001) alleviate these biases and also find lower convergence rates. The likelihood-based estimator employed in this paper represents an alternative approach to deal with finite-sample biases in dynamic panel data models without the requirement of stationarity assumptions.

[17]In the comparison between PWT 6.1 and PWT 6.2 with 67 regressors by

Ciccone and Jarocinski (2009), the average abs_diff is 0.08. If we redo their comparison with 34 regressors, the average abs_diff becomes 0.04. Given they use lower frequency data (one single 36-year period) than this paper (eight 5-year periods), as pointed out by Johnson et al. (2009) the results obtained using different revisions of the PWT are more robust with low frequency data. This represents a trade-off between robustness across PWT revisions and the number of observations available for estimation.

A Appendix

A.1 Computational Appendix

For the implementation of the empirical approaches described in the paper, we need to resort to the algorithms proposed in the literature because of the extremely large number of calculations required for obtaining the posterior mean and variance described in equations (6) and (7). This is because the number of potential regressors determines the number of models under consideration, for example, in our case, with $K = 35$ potential regressors, the number of models under consideration is 3.4x10^{10}. These algorithms carry out Bayesian Model Averaging without evaluating every possible model.

Concretely, for the BACE, BMA and BAMLE approaches we have made use of the Markov Chain Monte Carlo Model Composition (MC^3) algorithm proposed by Madigan and York (1995), which generates a stochastic process that moves through model space. The idea is to construct a Markov chain of models $\{M(t), t = 1, 2, ...\}$ with state space Ξ. If we simulate this Markov chain for $t = 1, ..., N$, then under certain regularity conditions, for any function $h(M_i)$ defined on Ξ, the average

$$\widehat{H} = \frac{1}{N}\sum_{t=1}^{N} h\left(M\left(t\right)\right)$$

converges with probability 1 to $E\left(h\left(M\right)\right)$ as $N \to \infty$. To compute (6) in this fashion, we set $h(M_i) = E(\theta|M_i, y)$.

To construct the Markov chain, we define a neighborhood $nbd(M)$ for each $M \in \Xi$ that consists of the model M itself and the set of models with either one variable more or one variable fewer than M. Then, a transition matrix $\mathbf{q}$ is defined by setting $\mathbf{q}(M \to M') = 0 \ \forall \ M' \notin nbd(M)$ and $\mathbf{q}(M \to M')$ constant for all $M' \in nbd(M)$. If the chain is currently in state M, then we proceed by drawing M' from $\mathbf{q}(M \rightarrow M')$. It is then accepted with probability

$$\min\left\{1, \frac{\Pr\left(M'|y\right)}{\Pr\left(M|y\right)}\right\}$$

Otherwise, the chain stays in state M.[18]

After some experimentation with generated data, we were able to verify the proper convergence properties of our GAUSS code which implements the described MC^3 algorithm.

A.2 Data Appendix

Table A1: Variable Definitions and Sources

Variable	Source	Definition
Dependent Variable	PWT 6.2	Growth of GDP per capita over 5-year periods (2000 US dollars at PPP)
Initial GDP	PWT 6.2	Logarithm of initial real GDP per capita (2000 US dollars at PPP)
Population Growth	PWT 6.2	Average growth rate of population
Population	PWT 6.2	Population in thousands of people
Trade Openness	PWT 6.2	Export plus imports as a share of GDP
Government Share	PWT 6.2	Government consumption as a share of GDP
Investment Price	PWT 6.2	Average investment price level
Labor Force	PWT 6.2	Ratio of workers to population
Consumption Share	PWT 6.2	Consumption as a share of GDP
Investment Share	PWT 6.2	Investment as a share of GDP
Urban Population	WDI 2005	Fraction of population living in urban areas
Population Density	WDI 2005	Population divided by land area
Life Expectancy	WDI 2005	Life expectancy at birth
Population under 15	Barro and Lee	Fraction of population younger than 15 years
Population over 65	Barro and Lee	Fraction of population older than 65 years
Primary Education	Barro and Lee	Stock of years of primary education
Secondary Education	Barro and Lee	Stock of years of secondary education
Political Rights	Freedom House	Index of political rights from 1 (highest) to 7
Civil Liberties	Freedom House	Index of civil liberties from 1 (highest) to 7
Malaria	Gallup et al.	Fraction of population in areas with malaria
Navigable Water	Gallup et al.	Fraction of land area near navigable water
Landlocked Country	Gallup et al.	Dummy for landlocked countries
Air Distance	Gallup et al.	Logarithm of minimum distance in km from New York, Rotterdam, or Tokyo
Tropical Area	Gallup et al.	Fraction of land area in geographical tropics

Table A1 - Continued

Variable	Source	Definition
Tropical Pop.	Gallup et al.	Fraction of population in geographical tropics
Land Area	Gallup et al.	Area in km^2
Independence	Gallup et al.	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 1989
Socialist	Gallup et al.	Dummy for countries under socialist rule for considerable time during 1950 to 1995
Climate	Gallup et al.	Fraction of land area with tropical climate
War Dummy	Barro and Lee	Dummy for countries that participated in external war between 1960 and 1990
SW Openness Index	Sachs, Warner	Index of trade openness from 1 (highest) to 0
Europe		Dummy for EU countries
Sub-Saharan Africa		Dummy for Sub-Saharan African countries
Latin America		Dummy for Latin American countries
East Asia		Dummy for East Asian countries

PWT 6.2 refers to Penn World Table version 6.2 and it can be found at http://pwt.econ.upenn.edu. WDI 2005 refers to World Development Indicators 2005. Data from Barro and Lee, Gallup et al., and Sachs and Warner are available at http://www.cid.harvard.edu/ciddata/ciddata.html. Finally, data from Freedom House can be downloaded from http://www.freedomhouse.org.

Table A2: List of Countries

Algeria	Indonesia	Peru
Argentina	Iran	Philippines
Australia	Ireland	Portugal
Austria	Israel	Rwanda
Belgium	Italy	Senegal
Benin	Jamaica	Singapore
Bolivia	Japan	South Africa
Brazil	Jordan	Spain
Cameroon	Kenya	Sri Lanka
Canada	Lesotho	Sweden
Chile	Malawi	Switzerland
China	Malaysia	Syria
Colombia	Mali	Thailand
Costa Rica	Mauritius	Togo
Denmark	Mexico	Trinidad & Tobago
Dominican Republic	Mozambique	Turkey
Ecuador	Nepal	Uganda
El Salvador	Netherlands	United Kingdom
Finland	New Zealand	United States
France	Nicaragua	Uruguay
Ghana	Niger	Venezuela
Greece	Norway	Zambia
Guatemala	Pakistan	Zimbabwe
Honduras	Panama	
India	Paraguay	

References

[1] Ahn, S. and P. Schmidt (1995) "Efficient Estimation of Models for Dynamic Panel Data" *Journal of Econometrics*, Vol. 68, pp. 5-27.

[2] Alvarez, J. and M. Arellano (2003) "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators" *Econometrica*, Vol. 71, No. 4. pp. 1121-1159.

[3] Arellano, M. (2003), "Panel Data Econometrics" Oxford University Press.

[4] Banerjee, M. and E. Frees (1997) "Influence Diagnostics for Linear Longitudinal Models" *Journal of the American Statistical Association*, Vol. 92, No. 439. pp. 999-1005.

[5] Barro, R. (1991) "Economic Growth in a Cross Section of Countries" *Quarterly Journal of Economics*, 106, 2, 407-43.

[6] Barro, R. and J.-W. Lee (1994) "Sources of Economic Growth" *Carnegie-Rochester Conference Series on Public Policy*, 40, 1-57.

[7] Bond, S., A. Hoeffler and J. Temple (2001) "GMM Estimation of Empirical Growth Models" *CEPR Discussion Papers*, No. 3048.

[8] Caselli, F., G. Esquivel and F. Lefort (1996) "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics" *Journal of Economic Growth*, 1, pp. 363-389.

[9] Chamberlain, G. (1984) "Panel Data" in Griliches, Z. and M. D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.

[10] Chen, H., A. Mirestean and C. Tsangarides (2009) "Limited Information Bayesian Model Averaging for Dynamic Panels with Short Time Periods" *IMF Working Paper* WP/09/74.

[11] Ciccone, A. and M. Jarocinski (2009) "Determinants of Economic Growth: Will Data Tell?" forthcoming in *American Economic Journal: Macroeconomics.*

[12] Durlauf, S., P. Johnson and J. Temple (2005) "Growth Econometrics" In P. Aghion and S.N. Durlauf, eds., *Handbook of Economic Growth*, Volume 1A, pp. 555-677, Amsterdam, North-Holland.

[13] Easterly, W. (1993) "How Much Do Distortions Affect Growth?" *Journal of Monetary Economics*, 32, 2, 187-212.

[14] Fernandez, C., E. Ley and M. Steel (2001a) "Model Uncertainty in Cross-Country Growth Regressions" *Journal of Applied Econometrics*, 16, pp. 563-576.

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

[16] Gallup, J., A. Mellinger and J. Sachs (2001) "Geography Datasets" Center for International Development (CID) at Harvard University.

[17] Islam, N. (1995) "Growth Empirics: A Panel Data Approach" *The Quarterly Journal of Economics*, Vol. 110(4), pp. 1127-70

[18] Johnson, S., W. Larson, C. Papageorgiou and A. Subramanian (2009) "Is Newer Better? Penn World Table Revision and Their Impact on Growth Estimates" NBER Working Paper No. 15455.

[19] Kass, R. and L. Wasserman (1995) "A Reference Bayesian Test for Nested Hypothesis with Large Samples" *Journal of the American Statistical Association*, Vol. 90, pp. 928-934.

[20] Koop, G. (2003) "Bayesian Econometrics" John Wiley and Sons.

[21] Krugman, P. (1991) "Increasing Returns and Economic Geography" *Journal of Political Economy*, Vol. 99 No. 3 pp. 483-499.

[22] Leamer, E. (1978) "Specification Searches" New York: John Wiley and Sons.

[23] Levine, R. and D. Renelt (1992) "A sensitivity Analysis of Cross-Country Growth Regressions" *American Economic Review*, 82, pp. 942-963.

[24] Ley, E. and M. Steel (2009) "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression" *Journal of Applied Econometrics*, Vol. 24, No. 4, pp. 651-674.

[25] Madigan, D. and J. York (1995) "Bayesian Graphical Models for Discrete Data" *International Statistical Review*, 63, pp. 215-232.

[26] Mirestean, A. and C. Tsangarides (2009) "Growth Determinants Revisited" *IMF Working Paper* WP/09/268.

[27] Mundlak, Y. (1978) "On the Pooling of Time Series and Cross Section Data" *Econometrica*, 46, pp. 69-85.

[28] Nickell, S. (1981) "Biases in Dynamic Models with Fixed Effects" *Econometrica*, Vol. 49, No. 6. pp. 1417-1426.

[29] Raftery, A. (1995) "Bayesian Model Selection in Social Research" *Sociological Methodology*, Vol. 25, pp. 111-163.

[30] Rodriguez, F. and Rodrik, D. (2000) "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence" *NBER Working Papers 7081*, National Bureau of Economic Research, Inc.

[31] Sachs, J. and Warner, A. (1997) "Natural Resource Abundance and Economic Growth" Center for International Development (CID) at Harvard University.

[32] Sala-i-Martin, X. (1997) "I Just Ran Two Million Regressions" *American Economic Review*, Vol. 87, No. 2. pp. 178-183.

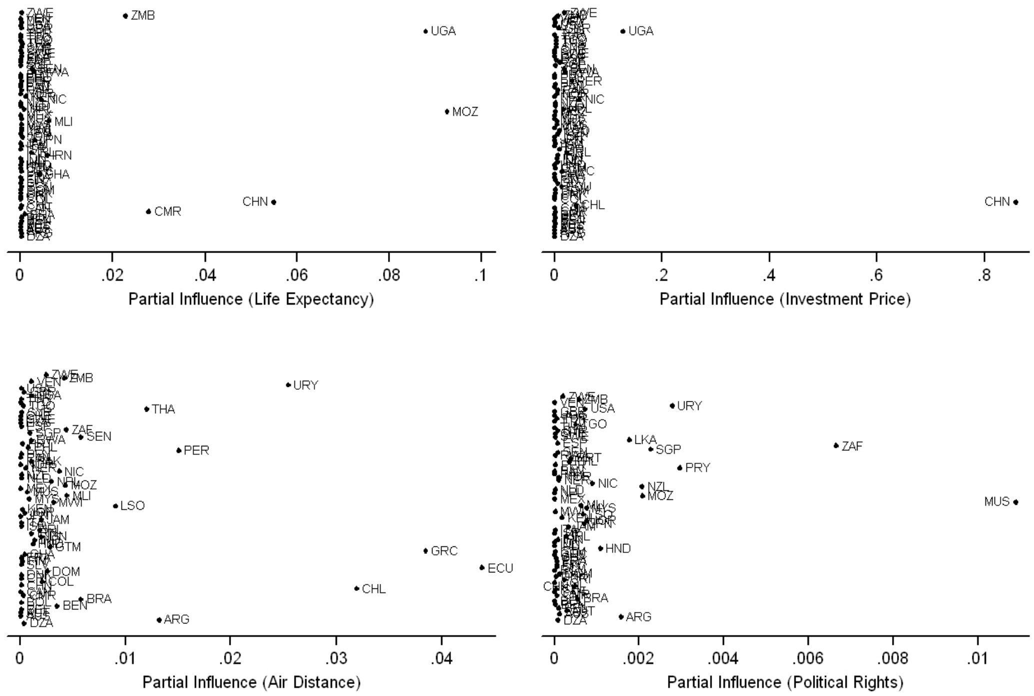
[33] Sala-i-Martin, X., G. Doppelhofer and R. Miller (2004) "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach" *American Economic Review*, Vol. 94, No. 4. pp. 813-835.

[34] Schwarz, G. (1978) "Estimating the Dimension of a Model" *The Annals of Statistics*, Vol. 6, No. 2. pp. 461-464.

[35] Temple, J. (1998) "Robustness Tests of the Augmented Solow Model" *Journal of Applied Econometrics*, Vol. 13, No. 4. pp. 361-375.

Figures

Figure 1: Dotplots of the Partial Influence Measure*



*In this figure we plot the Banerjee and Frees' influence statistics. For all the four variables the statistic is computed N times, one for each country. Under the null of no influence the statistics are approximately distributed according to a χ^2_1 distribution. Since the 5% critical value is 3.84 we conclude that there are not "influential countries".

Tables

Table 1: Posterior Inclusion Probability of the Regressors

Variable	ξ Fixed				ξ Random			
	$m = 5$		$m = K/2$		$m = 5$		$m = K/2$	
	SDM (1)	FLS (2)	SDM (3)	FLS (4)	SDM (5)	FLS (6)	SDM (7)	FLS (8)
Initial GDP	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Population	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Population under 15	0.950	0.961	0.937	0.953	0.953	0.965	0.949	0.963
Investment Share	0.826	0.847	0.783	0.835	0.822	0.841	0.816	0.843
Urban Population	0.651	0.392	0.781	0.596	0.608	0.358	0.638	0.387
Consumption Share	0.305	0.100	0.682	0.229	0.303	0.088	0.351	0.099
Trade Opennes	0.287	0.106	0.656	0.218	0.289	0.094	0.336	0.103
Government Share	0.237	0.064	0.549	0.173	0.231	0.058	0.273	0.068
Investment Price	0.222	0.088	0.376	0.176	0.206	0.083	0.229	0.092
Population Density	0.031	0.013	0.061	0.024	0.029	0.011	0.033	0.013
Labor Force	0.029	0.013	0.064	0.022	0.028	0.010	0.033	0.012
Primary Education	0.026	0.010	0.061	0.023	0.026	0.009	0.030	0.010
Civil Liberties	0.023	0.007	0.053	0.017	0.022	0.006	0.025	0.008
Population Growth	0.018	0.005	0.050	0.013	0.019	0.005	0.022	0.005
Life Expectancy	0.018	0.006	0.051	0.013	0.019	0.005	0.023	0.006
Malaria	0.020	0.005	0.043	0.014	0.018	0.006	0.021	0.006
Population over 65	0.017	0.005	0.044	0.013	0.018	0.004	0.021	0.006
Secondary Education	0.017	0.005	0.046	0.012	0.017	0.005	0.020	0.005
Political Rights	0.016	0.005	0.044	0.012	0.016	0.004	0.020	0.005
Prior Mean Model Size	5	5	9	9	5	5	9	9
Post. Mean Model Size	5.69	4.63	7.28	5.34	5.62	4.55	5.83	4.63

Column heading SDM refers to the BACE-SDM Approach in a panel data context and column heading FLS refers to BMA-FLS approach in a panel data context. The purpose of this table is to illustrate that the effects of different prior assumptions on model size are much less severe with the hierarchical priors (ξ Random) proposed by Ley and Steel (2009).

Table 2: Panel SDM-FLS Approaches Results

Variable	Posterior Inclusion Probability		Posterior Mean		Posterior Standard Deviation	
	SDM	FLS	SDM	FLS	SDM	FLS
Initial GDP	1.000	1.000	-0.271	-0.265	0.029	0.030
Population	1.000	1.000	0.918	0.905	0.176	0.176
Population under 15	0.953	0.965	-1.122	-1.183	0.287	0.279
Investment Share	0.822	0.841	0.343	0.351	0.097	0.095
Urban Population	0.608	0.358	-0.426	-0.433	0.147	0.147
Consumption Share	0.303	0.088	-0.210	-0.202	0.068	0.091
Trade Opennes	0.289	0.094	0.102	0.100	0.028	0.046
Government Share	0.231	0.058	-0.336	-0.315	0.140	0.149
Investment Price	0.206	0.083	-0.031	-0.033	0.014	0.014
Population Density	0.029	0.011	0.042	0.063	0.054	0.057
Labor Force	0.028	0.010	0.225	0.363	0.415	0.477
Primary Education	0.026	0.009	-0.169	-0.194	0.179	0.186
Civil Liberties	0.022	0.006	-0.044	-0.047	0.060	0.060
Population Growth	0.019	0.005	-0.488	-0.317	1.156	1.091
Life Expectancy	0.019	0.005	0.063	-0.011	0.241	0.250
Malaria	0.018	0.006	0.010	0.013	0.024	0.026
Population over 65	0.018	0.004	-0.220	-0.200	0.824	0.801
Secondary Education	0.017	0.005	-0.051	-0.034	0.186	0.191
Political Rights	0.016	0.004	-0.009	-0.004	0.048	0.049

Column heading SDM refers to the BACE-SDM Approach in a panel data context, and column heading FLS refers to BMA-FLS approach in a panel data context. All results presented in this Table are based on prior assumptions $m = 5$ and ξ Random. The results with $m = K/2$ are not presented here for the sake of brevity, but they were practically identical.

Table 3: BAMLE Approach Results

Variable	Posterior Inclusion Probability	Posterior Mean	Posterior Standard Deviation
Initial GDP	1.000	-0.033	0.035
Life Expectancy	1.000	0.145	0.287
Investment Price	0.863	-0.049	0.015
Air Distance	0.759	-0.962	0.381
Political Rights	0.722	-0.053	0.013
Population Growth	0.688	-1.082	1.081
Urban Population	0.650	-0.475	0.163
Population	0.639	0.602	0.201
Trade Openness	0.467	0.056	0.020
Landlocked Country	0.320	-0.346	0.359
Investment Share	0.238	0.271	0.105
Civil Liberties	0.176	0.048	0.017
Government Share	0.161	-0.160	0.148
Latin America	0.147	0.038	0.015
Population Density	0.087	-0.014	0.081
East Asia	0.073	-0.012	0.006
Consumption Share	0.057	0.036	0.062
Navigable Water	0.057	0.043	0.026
Europe	0.052	-0.036	0.018
Tropical Area	0.034	-0.252	0.201
Sub-Saharan Africa	0.029	0.027	0.021
Climate	0.028	-0.014	0.013
Primary Education	0.028	0.024	0.022
Tropical Pop.	0.025	-0.144	0.212
Labor Force	0.023	0.028	0.394
Population over 65	0.022	-0.012	0.018
SW Openness Index	0.018	-0.033	0.069
Land Area	0.017	0.021	0.056
War Dummy	0.017	0.001	0.019
Population under 15	0.017	0.010	0.012
Secondary Education	0.017	-0.008	0.016
Independence	0.016	-0.002	0.015
Socialist	0.016	-0.009	0.013
Malaria	0.013	0.001	0.012

This Table presents the results of applying the BAMLE methodology described in Section 3. The combination of a correlated random-effects maximum likelihood estimator with Bayesian Model Averaging techniques allows simultaneously addressing endogeneity biases and model uncertainty. All results in this Table are based on prior assumptions $m = 5$ and ξ Random.

Table 4: Sensitivity Analysis Results

Panel A: PWT 6.2 versus PWT 6.1

	Fixed Effects		No Fixed Effects	
	K=19	K=10	K=19	K=10
Average abs_diff PIP	0.077	0.037	0.092	0.034
Median abs_diff PIP	0.019	0.011	0.032	0.019

Panel B: PWT 6.2 versus PWT 6.3

	Fixed Effects		No Fixed Effects	
	K=19	K=10	K=19	K=10
Average abs_diff PIP	0.051	0.028	0.053	0.027
Median abs_diff PIP	0.005	0.004	0.029	0.014

Panel C: PWT 6.2 versus WDI 2005

	Fixed Effects		No Fixed Effects	
	K=19	K=10	K=19	K=10
Average abs_diff PIP	0.179	0.044	0.206	0.096
Median abs_diff PIP	0.075	0.011	0.018	0.012

This Table presents measures of sensitivity of the results when using different sources of international GDP data. In particular, following Ciccone and Jarocinski (2009) it reports the average and median absolute values of the difference between Posterior Inclusion Probabilities (PIP) for all the variables (lower values indicate smaller sensitivity). For comparison purposes we consider different numbers of candidate regressors, K=19 or K=10, and we allow for country-specific effects or not. The sample comprises 73 countries and eight 5-year periods over 1960-2000 in Panels A and B. Given WDI 2005 data availability, in Panel C the sample period is 1975-2000.