Unraveling the fortunes of the fortunate: An Iterative Bayesian Model Averaging (IBMA) approach

Theo S. Eicher [a,b,*], Chris Papageorgiou [c,d], Oliver Roehn [b]

[a] Department of Economics, University of Washington, WA 98195-3330, USA
[b] Ifo Institute for Economic Research at the University of Munich, Germany
[c] Research Department, International Monetary Fund, USA
[d] Department of Economics, Louisiana State University, USA

Abstract

We investigate country heterogeneity in cross-country growth regressions. In contrast to the previous literature that focuses on low-income countries, this study also highlights growth determinants in high-income (OECD) countries. We introduce Iterative Bayesian Model Averaging (IBMA) to address not only potential parameter heterogeneity, but also the model uncertainty inherent in growth regressions. IBMA is essential to our estimation because the simultaneous consideration of model uncertainty and parameter heterogeneity in standard growth regressions increases the number of candidate regressors beyond the processing capacity of ordinary BMA algorithms.

Our analysis generates three results that strongly support different dimensions of parameter heterogeneity. First, while a large number of regressors can be identified as growth determinants in Non-OECD countries, the same regressors are irrelevant for OECD countries. Second, Non-OECD countries and the global sample feature only a handful of common growth determinants. Third, and most devastatingly, the long list of variables included in popular cross-country datasets does not contain regressors that begin to satisfactorily characterize the basic growth determinants in OECD countries.
© 2007 Elsevier Inc. All rights reserved.

* Corresponding author. Address: Department of Economics, University of Washington, WA 98195-3330, USA.

   E-mail address: te@u.washington.edu (T.S. Eicher).

1. Introduction

Over the last two decades there has been a massive effort to use cross-country datasets to identify key determinants of economic growth. Much of this empirical investigation has been based on the implicit assumption of homogeneity across countries, which led to a search for *global* determinants of growth. However, the assumption of homogeneity in cross-country growth regressions has been criticized repeatedly (see e.g. Temple, 2000; Durlauf et al., 2005). In general, this objection applies to any socioeconomic dataset but the assumption of a common underlying data generating process seems particularly inappropriate when analyzing such complex entities as countries (Brock and Durlauf, 2001).

The mounting evidence against "country homogeneity" has given rise to a literature investigating growth patterns in groups of countries that share common characteristics. This branch of research focuses either on particular variables (e.g., initial GDP) or particular regions (Africa, Latin America) that distinguish subsamples.[1] In this paper we revisit the issue of country heterogeneity but from a perspective that has been largely ignored by the empirical growth literature. We focus on identifying growth determinants in high-income (OECD) countries, to understand the structures that drive the riches in industrialized countries that developing nations attempt to emulate. In essence, our goal is to understand the driving forces behind sustained economic success, with the assumption that such successful growth paths are determined by a unique set of variables. Eicher and Leukert (2006) previously explored parameter heterogeneity among OECD and Non-OECD countries, but did not account for model uncertainty or a large number of potential regressors.

Our estimation approach includes both parameter heterogeneity, to allow countries to represent diverse objects, and model uncertainty, to account for the fact that economists do not know the single "true" growth model. More specifically, we utilize Bayesian Model Averaging (BMA) to address model uncertainty and expand the methodology to integrate structures that allow for the examination of parameter heterogeneity. Simultaneous consideration of model uncertainty and parameter heterogeneity has previously been computationally prohibitive, as it exceeded the computational limits of existing model averaging algorithms. This is due to the large numbers of candidate regressors that emerge from the long list of potential growth regressors and relevant interaction terms that are required to test for parameter heterogeneity. To resolve the computation limitations we employ an innovative modification of BMA called Iterative Bayesian Model Averaging (IBMA) developed by Yeung et al. (2005) for genomics applications. The key intuition of IBMA is that it applies traditional BMA iteratively on a reduced set of variables. Each iteration contains a set of variables that is sufficiently small to be processed by existing algorithms.

---

[1] See e.g. Easterly and Levine (1997), Brock and Durlauf (2001) and Masanjala and Papageorgiou (2007a,b).

Iterations continue until the complete set of candidate regressors has been processed at least once.

We obtain three key results that highlight different dimensions of country heterogeneity. First, of the large number of regressors that are effective in the global sample, only about half are also effective in the Non-OECD sample. This is surprising, since the large number of countries in the Non-OECD sample were thought to be providing most of the explanatory power for the global results. Secondly, our analysis shows that in Non-OECD countries new regressors become highly effective that were ineffective in the global sample. Many of these newly effective variables are highly intuitive, for example the primary export share, black market premium, average population age. Third, the OECD subsample shares few regressors with the global sample (6 out of 20); this leads us to conclude that the particular dataset does not contain the variables that identify determinants of growth of the fortunate in the past 30 years. There are also stark difference between OECD and the Non-OECD sample where only half of the variables overlap.

The rest of the paper is organized as follows. Section 2 presents a summary of BMA and IBMA methodologies used in our econometric estimation. Section 3 discusses the cross-country dataset used, and presents the benchmark regression specification based on which we perform IBMA. This section also presents and examines the estimation results. Section 4 presents robustness analyses of our results to alternative modifications of the sampler used by IBMA. Section 5 concludes and offers directions for future research.

2. Estimation methodology

The basic idea behind model averaging is to estimate the distribution of unknown parameters of interest across different models. The fundamental principle of model averaging is to treat models and related parameters as unobservable, and to estimate their distributions based on the observable data. In contrast to classical estimation, model averaging copes with model uncertainty by allowing for all possible models to be considered, which consequently reduces the biases of parameters.

Leamer (1978) first emphasized that the uncertainty inherent in competing theories should be accounted for in the empirical strategy. Levine and Renelt (1992) examine the robustness of cross-country growth determinants using Leamer (1983) extreme bounds analysis. They show that the conclusions as to which regressors represent robust growth determinants depends on the researcher's test criteria. Extreme bound analysis has since been shown to be excessively strict, selecting too few "effective" regressors (see, Sala-i-Martin, 1997 for a criticism of this approach relevant to growth regressions).

An additional drawback of extreme bound analysis has been the absence of a formal structure to manage the large number of possible models. Levine and Renelt (1992) choose to reduce the set of models to be examined by always including Initial Income, Investment Rates, Secondary School Enrollment Rate, and Population Growth Rate in each regression. Sala-i-Martin (1997) used the same method, but he chose to always retain Initial Income, Investment Rates and Life Expectancy. Fixing the number of regressors that must appear in each regression has a direct effect on the size of the estimated coefficients (see Leon-Gonzalez and Montolio, 2003) and it limits the number of models that are explored.

Since the first approaches to model uncertainty, a consensus has formed to apply Bayesian techniques to account for model uncertainty (see e.g. Fernandez et al., 2001a,b; Brock

and Durlauf, 2001; Sala-i-Martin et al., 2004; and Masanjala and Papageorgiou, 2007a,b). Model averaging strategies asks the researcher to specify candidate regressors that are clearly linked to distinct and specific theories. Bayesian Model Averaging then allows for any subset of regressors to appear in a given model. This technique was first developed by Multon (1991), and Palm and Zellner (1992), but computational issues initially hampered its implementation.[2] Since our methodology is based on BMA, we provide a brief overview of the method.

2.1. Bayesian model averaging

Bayesian Model Averaging (BMA) accounts for model uncertainty by averaging over all possible models, where each model's weight is given by its posterior model probability. The statistical foundation for BMA is documented extensively in excellent introductions by Raftery (1995) and Hoeting et al. (1999). Raftery (1995) and Raftery et al. (1997), followed by many others, have shown that BMA provides improved out-of-sample predictive performance compared to predictions that are conditioned on any one model.

We restrict ourselves to highlighting the crucial intuition behind the methodology and then provide an explanation of the specific approach that we implemented together with the methodological innovations. In typical cross-country growth regressions, model uncertainty arises due to the fact that the researchers must choose between regressors that are associated with competing theories. With k possible variables in a linear regression model, BMA potentially considers the entire model space of 2^k regression models. The posterior probability that BMA assigns is simply the conditional probability after all relevant data has been taken into account. Posterior probabilities are calculated using Bayes' theorem, utilizing the researcher-specified prior probability and the likelihood function.

Formally, consider n independent replications from a linear regression model where the dependent variable is per capita GDP growth, y, is regressed on an intercept, α, and candidate regressors chosen from a set of k variables in a design matrix Z of dimension $n \times k$. Assume that the rank of the matrix of regressors is $r(\iota_n{:}Z) = k + 1$, where ι_n is an n-dimensional vector of ones. Further define β as the full k-dimensional vector of regression coefficients. Now suppose we have an $n \times k_j$ submatrix of variables in Z denoted by Z_j. Then denote by M_j the model with regressors grouped in Z_j, such that

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

where $\beta_j \in \Re^{k_j} (0 \leqslant k_j \leqslant k)$ groups regression coefficients corresponding to the submatrix Z_j. The exclusion of any given regressor in a particular model implies that the corresponding element in β is zero. $\sigma \in \Re_+$ is a scale parameter and ε follows an n-dimensional normal distribution with zero mean and identity covariance matrix.

Since Bayesian Model Averaging allows for any subset of variables in Z to appear in any model M_j, thus there are 2^k possible sampling models. BMA specifies that the posterior inclusion probability of any given parameter of interest is the weighted posterior distribution of that quantity under each of the models. The specific weights are provided by each model's posterior model probability. The posterior inclusion probability can then be

---

[2] For further discussions on BMA and its potential uses see Draper (1995), Raftery et al. (1997) and Hoeting et al. (1999).

expressed as the weighted sum of the posterior probabilities of all models that contain the regressor of interest

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

The posterior model probability itself is given by

$$P(M_j|y) = \frac{l_y(M_j)p_j}{\sum_{h=1}^{2^k} l_y(M_h)p_h}, \tag{3}$$

where $l_y(M_j)$, is the marginal likelihood of model M_j that is given by

$$l_y(M_j) = \int p(y|\alpha, \beta_j, \sigma, M_j)p(\alpha, \sigma)p(\beta_j|\alpha, \sigma, M_j)\mathrm{d}\alpha\mathrm{d}\beta_j\mathrm{d}\sigma. \tag{4}$$

The sampled model corresponding to Eq. (1) is given by $p(y|\alpha, \beta_j, \sigma, M_j)$, and the priors for the intercept and the regressors are $p(\alpha, \sigma)$ and $p(\beta_j|\alpha, \sigma, M_j)$, respectively. We will define the priors below.

The implementation of Bayesian Model Averaging is subject to three challenges. First, the number of models that must be estimated increases with the number of regressors at the rate of 2^k. As a result, the number of summation entries in Eqs. (2) and (3) can be enormous; a primary aim of BMA research has been to obtain efficient samplers that avoid exhaustive sampling. Such intensive calculations quickly become infeasible as 30 candidate variables imply over 1 billion candidate models. Second, the computation and evaluation of the integrals implicit in Eq. (4) may be difficult because they may not exist in closed form. In that case numerical solutions of the integral can further burden estimation efficiency. Third, the choice of the prior distribution specification is always contentious in Bayesian analysis. BMA requires the specification of two types of priors: (a) prior model probabilities, $p(M_K)$, and (b) prior parameter distribution $p(\theta_K, M_K)$.

With respect to the prior model probabilities we follow the common practice in the literature and assume a uniform distribution over the model space, which expresses each model as equally likely. It follows that the prior model probability is 2^{-k}, which renders the prior probability of including any given candidate regressor equal to 0.5 (see e.g., Raftery et al., 1997; and Fernandez et al., 2001a,b).[3]

The decision on the prior structure for the individual regressors is a potentially divisive issue. BMA requires the researcher to inject priors into the analysis, however these prior can be so diffuse that clear parallels to frequentist inference can be established. Extensive work has been conducted on the appropriate prior structure to obtain either data dependent priors (Raftery et al., 1997), "automatic" priors (Fernandez et al., 2001b), or the Unit Information Prior (UIP). Eicher et al. (2007) systematically study the effects of model and regressor priors on predictive performance within a BMA framework to highlight the importance of a prior benchmark. Their software allows researchers to identify the appropriate prior structure for a given dataset.

---

[3] Mitchell and Beauchamp (1988) discuss the possibility of alternative model weights and Sala-i-Martin et al. (2004) argue forcefully in favor of greater weights on smaller models. Brock et al. (2003) suggest a tree structure to take into account similarities among regressors.

In our choice regarding the priors on the parameters space we follow Raftery (1995) and impose the diffuse UIP. The UIP can be derived from frequentist statistical principles (Kass and Wasserman, 1995), and it is seen as a conservative prior that is sufficiently spread out over the relevant parameter values and reasonably flat over the area where the likelihood is substantial. Specifically, it is a multivariate normal prior with mean at the maximum likelihood estimate and variance equal to the expected information matrix for one observation (Raftery, 1999). It is also a special case of the preferred Fernandez et al. (2001b) priors and it is closely related to the prior structure in Sala-i-Martin et al. (2004). The advantage of the UIP is that it allows for a simple approximation of the marginal likelihood with the Bayesian Information Criterion (BIC). The BIC approximation is viewed as conservative fitness measure to evaluate model performance. If anything, BIC is biased against finding an effect of a given regressor (i.e. it favors the null hypothesis $\beta = 0$).[4]

The one crucial departure from previous applications of model averaging in economics is our sampling and estimation methodology. Fernandez et al. (2001a,b) use the Markov Chain Monte Carlo Model Composition (MC^3) sampling algorithm developed by Madigan and York (1995) to search the model space, while Sala-i-Martin et al. (2004) use a ''stratified'' Coinflip sampler. MC^3 is a technique that allows for sampling of complex high dimensional distributions as it simulates a random walk across the search space to converge at a stationary posterior distribution. The MC^3 distribution of the sampled draws depends on the last value drawn. In contrast, the stratified Coinflip sampler samples one set of regressions using the prior probability sampling weights and then uses the approximate posterior inclusion probabilities calculated from those regressions for the subsequent sampling probabilities.

Given that MC^3's computational limit was no more than 60 candidate regressors,[5] the Coinflip sampler had the advantage of handling more candidate regressors. However, the larger the search space the more difficult was for Coinflip sampler to converge. For example, in some BMA experiments we run with more than 70 candidate regressors there was no (or unacceptably slow) converge simply because the number of models becomes too large.

Our method follows Raftery (1995) who established that the UIP allows for a Laplace approximation of the marginal likelihood and thus renders a search across the entire model space obsolete. To further simplify the computational demands Raftery (1995) suggest the *Leaps And Bounds All Subsets Regression Algorithm* of Furnival and Wilson (1974) to reduce the candidate model space further.[6] The Leaps algorithm performs an exhaustive search for the best subsets of candidate variables for predicting the dependent variable in linear regression; it returns a specified number of best models for each model size.[7] Generally, the qualitative differences based on the different samplers are small but not negligible. Computationally, the Leaps sampler is by far the most efficient. This efficiency is

---

[4] See e.g. Raftery (1995). For a more detailed discussion of the UIP and BIC, see Raftery (1999) and the discussion in Hoeting et al. (1999).

[5] At least until very recently. We have just discovered that the work of Ley and Steel (in this issue) extends the computational bound of MC^3 to 104 regressors. We discuss this development in the end of this section.

[6] See e.g. Raftery (1995) and Volinsky et al. (1997).

[7] Software to implement the Raftery method has been freely available since 1994 at Statlib (http://lib.stat.cmu.edu).

crucial to handle the large number of models as we tackle model uncertainty and parameter heterogeneity by interacting the global variables with regional dummies, which substantially increases the size of candidate regressors.

2.2. Iterative bayesian model averaging

The computational limit of the Raftery (1995) BMA algorithm (*bicreg*) is 54 candidate regressors. To address parameter heterogeneity, the interaction of regressors increases the domain of regressors from 41 to a possible 82, which implies 4 septillion (100 billion × 4 trillion) models. In addition, the simple act of interacting variables in a small dataset may lead the number of regressors to exceed the number of observation, such that the design matrix is no longer of full column rank.

To overcome these problems we introduce the Iterative BMA (IBMA) algorithm to economics that was initially proposed for a genome application by Yeung et al. (2005). Specifically, they introduced IBMA to select a small number of relevant genes for accurate medical diagnoses from a pool of about 5000(!) genes. Our application is simpler. After interacting our 41 regressors with an OECD treatment dummy and eliminating interaction terms that are perfectly collinear or have less than 2 observations, this leaves us with 77 candidate regressors (see the data discussion below).

The key intuition of IBMA is that it applies traditional BMA iteratively on a reduced set of variables, z, which is small enough to be processed by traditional BMA. We define z as the *regressor window*. For our application we choose a default size $z = 41$ and check for robustness below. After sorting the candidate regressors by their bivariate correlations with the dependent variable, they are added to the regressor window. After the first z regressors have been processed by the first BMA run, q variables whose posterior probabilities do not exceed a predetermined inclusion threshold (1% by default) are removed from the regressor window and q unprocessed candidate regressors are added. BMA is then applied again until all regressors have been considered.

There are some caveats that must be highlighted as the set of candidate regressors expands. One limiting factor for IBMA is related to the regressor window size.While models of size n are theoretically possible, IBMA cannot evaluate posteriors for models that exceed size z. Hence the procedure cannot lay claim to having examined the entire model space – which introduces possible inaccuracies if high quality models happen to be larger than z. In our robustness section we find that variations in z in IBMA do not alter our qualitative results in the growth dataset.

Although we provided this caveat, we can offer evidence that any concerns that z may not cover the relevant model size are unlikely to be applicable in cross-country growth regressions. Sala-i-Martin (1997) and Sala-i-Martin et al. (2004) argue forcefully that the expected model size should not exceed 7 regressors. Prior work by Levine and Renelt (1992), Sala-i-Martin (1997), FLS and Sala-i-Martin et al. (2004) never generated models with more than 18 potentially relevant regressors. Hence it is unlikely that high quality models in cross-country growth regressions contain more than 48 regressors.

New work by Ley and Steel (in this issue) extends MC[3] to potentially handle up to 104 regressors without the iterating procedure employed in our algorithm. The advantage being that the entire model space, including models up to 104 regressors can actually be considered. This also implies that the prior model size increases to perhaps an implausibly large number of regressor, however. It remains to be seen how accurate and time intensive the

new MC[3] method generates convergence. Previous work using MCMC methods, particularly in applications with growth datasets, revealed that increasing the number of regressors (which of course increases the model space exponentially) resulted in considerable increase in computation time. Alternatively, IBMA is not limited to the number of candidate regressors and processes the data with stunning efficiency. It also allows the researcher to avoid having the prior model size increase linearly with the number of candidate regressors. Further research is necessary to examine how the three existing approaches to considering large model spaces (IBMA, modified MC[3] and BACE) compete in terms of efficiency and predictive performance. The unique advantage of IBMA over the other two approaches, at least to date, is that it is capable of considering applications like ours where the number or observations happen to be less than the number of potential regressors.

3. Estimation

3.1. The data

For our analysis we adopt the FLS dataset. It is comprised of 41 variables and 72 countries of which 23 are OECD countries. In addition, we add a dummy variable to identify OECD countries. The dataset is a subset of the Sala-i-Martin (1997) dataset; it includes all variables that have previously been flagged as robustly related to growth and that do not entail a loss of observations. We choose the FLS dataset for several reasons. First, the dataset contains variables that proxy for a broad set of competing growth theories, such as human capital, institutional quality, religion, economic policy and geography. Hence, the dataset reflects the theory uncertainty inherent in growth econometrics that has been highlighted by Brock and Durlauf (2001). Second, the majority of the variables are measured at the beginning of the period or as close as possible to it, which reduces possible endogeneity problems that can potentially impact cross-country growth regression analyses. Finally, by choosing the same dataset as FLS we have a natural benchmark and reference point for our analysis.

Table A1 in the Appendix provides summary statistics for the global, OECD, and Non-OECD samples. The high income OECD countries grew on average almost twice as fast as the rest of the world over the period 1960–1992 (3% versus 1.7%). A first look at the data reveals some major initial advantages OECD countries possessed over the rest of the world. In 1960, initial GDP was about four times greater, life expectancy was 16 years greater and primary schooling was 28% higher in the OECD sample as compared to the Non-OECD sample. OECD economies also had effectively better institutions scoring higher on civil liberties, the rule of law and political rights[8], while ethnolinguistic fractionalization was twice as high in Non-OECD countries.

3.2. Model specification

To examine the possibility of parameter heterogeneity, we examine whether the data generating process for the global sample is different from the data generating process of

---

[8] Note that Civil Liberties and Political Rights are measured "backwards," i.e. larger values imply fewer civil liberties and political rights.

the OECD sample.[9] To model parameter heterogeneity we follow the approach suggested by Brock and Durlauf (2001) and Brock et al. (2003) and treat parameter heterogeneity as a variable inclusion problem. It follows then that we can understand parameter heterogeneity as a special case of model uncertainty. We therefore modify the global equation in (1) and estimate the standard interaction model in empirical work of the following form:

$$y = \alpha \iota_n + Z_j \beta_{1,j} + I X_j \beta_{2,j} + \sigma \varepsilon, \tag{5}$$

where I is an indicator variable that equals 1 if the country is an OECD member and 0 otherwise. Z is the $n \times k$ matrix of the regressors and X is a sub-matrix of Z that excludes all variables that are either perfectly collinear in the OECD sample[10] or not relevant for the OECD sample due to negligible sub-sample variation.[11] In our case with OECD interactions, the resulting model features 77 candidate regressors and 72 observations, which renders traditional BMA infeasible and leads us to implement the IBMA algorithm discussed above. The direct merit of the interaction model compared to subsample regressions is that the full information from the entire dataset is used to derive results.

Regression Eq. (5) can be interpreted as providing estimates for the control group, β_{1i}, which is in our case the sample of Non-OECD countries. It also provides the marginal effect experienced by the treatment group, β_{2i}, which are the OECD countries in our case. The actual impact of the X regressors for which we want to establish parameter heterogeneity can then be obtained by comparing the Non-OECD effect given by the posterior means of β_{1i} with the effect in OECD countries that is given by the composite means of $\tilde{\beta}_i = \beta_{1i} + \beta_{2i}$.[12] Note that the definition of the composite $\tilde{\beta}_i$ carries an important implication: If the Non-OECD effect, $\beta_{1,i}$, is observed to be significantly different from zero and the OECD effect is found to be insignificant, it implies either that the marginal estimate of the treatment group, β_{2i}, is estimated with great noise (e.g., with a high variance) to wash out any significance of the composite, or that the treatment effect is indeed quite tightly estimated, but of the opposite sign as, β_{1i}, rendering the composite $\tilde{\beta}_i$ close to zero.

At this point it is important that the basic iterative routine suggested by Yeung et al. (2005) must be modified to assure that, β_{1i} and β_{2i} can appear in the same regression. Two cases are possible. In the first case, β_{1i} is included in a regressor window but the interaction is not (perhaps because its initial bivariate correlation was low). The rotation of each variable that is not in the initial regressor window does assure that β_{1i} and β_{2i} are in the final regressor window if they are both significant. In the second case, the initial regressor window includes β_{2i} but not β_{1i}, and the interaction alone is not significant. In this case the interaction will be rotated out of the regressor window and β_{2i} will never have the chance to actually interact with β_{1i}. This case requires a modification of the Yeung et al. (2005) procedure. In particular, we allow for two rotations in our version of IBMA.

---

[9] Theoretical underpinings for parameter heterogeneity are based on thresholds as in Azariadis and Drazen (1990), or on fully specified models of nonlinearities as in Galor and Weil (2000), Lucas (2002) and Galor and Moav (2002).

[10] The presence of multicollinearity exacerbates the problem of distinguishing between interaction terms that represent parameter heterogeneity and terms that are simply feature highly correlated with important interactions. This problem is neither unique to our issue at hand (OECD interaction), or IBMA.

[11] Excluded interactions are: Africa dummy, French Colony dummy, Fraction Hindu, Latin American dummy, Spanish Colony dummy, Fraction Confucian and Fraction Buddha.

[12] The composite variance is given by $\mathrm{var}(\tilde{\beta}_i) = \mathrm{var}(\beta_{1i}) + \mathrm{var}(\beta_{2i}) + 2\mathrm{cov}(\beta_{1i}, \beta_{2i})$.

The first rotation (as suggested by Yeung et al., 2005) assures that all regressors that were not included in the initial regressor window will have a chance to be considered. The second rotation iterates all regressors that have been discarded from regressor windows in the first rotation (to make room for new regressors) once more through the window. This assures that even in the second case, an initially discarded interaction term will have the opportunity to eventually rejoin the global variable in a regressor window, if significant.[13]

Further considerations to assure that variables have been given due chance to exhibit their true interaction significance in IBMA led us to examine the final regressor window to see how many global terms were observed without interaction terms. As a robustness exercise, we executed final iterations that added interaction terms to match all significant global regressors whose interaction terms did not appear in the final regressor window.[14]

Our empirical strategy is to start by establishing the global BMA benchmark, in Table 1. Here we initially examine the potential effectiveness of variables without any interactions specified in Eq. (5). Then we examine potential evidence for parameter heterogeneity. Finally we will examine robustness and compare different regressor window sizes in IBMA where we iterate until all covariates have been processed and the interaction terms are all included in the last iteration.

3.3. Results

Table 1 presents our baseline results applying IBMA to examine model uncertainty and parameter heterogeneity in the FLS dataset. In particular, Table 1 presents the coefficient posterior means, posterior standard deviation and the ratio of the absolute value of the former to the later, for the Global and Interaction specifications. The value of the absolute value of the posterior mean to standard deviation ratio (post. mean/s.d.) is used as a measure for identifying variable effectiveness in our growth regression exercises. While the analysis of posterior inclusion probability speaks only to the probability of a candidate repressor's inclusion in the most effective models, we chose to emphasize the post. mean/s.d. ratio to better tie economic and statistical significance. Raftery (1995) suggested that for a variable to be considered as effective the posterior inclusion probability must exceed 50%; which is roughly equivalent of requiring a ratio of mean/s.d. = 1, which implies in frequentist statistics that the regressors improves the power of the regression. Hence, while Raftery's (1995) interpretation for BMA would imply a threshold value of the mean/s.d. ratio of about 1, we decided to be more stringent and set the threshold value equal to 1.3, which is roughly equivalent to a 90% confidence interval in frequentist hypothesis testing. We recognize that there is no consensus in the BMA literature about this threshold, but argue that our main results hold when this threshold is adjusted upwards or downwards.

The results for the interaction model are obtained by using IBMA with a regressor window of size $z = 41$. The choice of the regressor window size is natural in that it is directly

---

[13] The Yeung et al. (2005) algorithm also suffers from the fact that it guarantees that the covariate with the lowest bivariate correlation is included in the final regressor window and hence in the final result. By adding regressors in the second rotation in inverse bivariate correlation order, we also improve on this design flaw.

[14] In additional robustness analysis, we also added one global regressor that was associated with one highly significant interaction term (Standard Deviation of the Black Market Premium) into the regressor window. This variable was found to be important for OECD but not robust across different windows considered. Our remaining results were unaffected.

Table 1
Effective growth determinants in global and interaction models

	Global			Interaction		
	Posterior mean	Posterior s.d.	Posterior mean/s.d. ratio	Posterior mean	Posterior s.d.	Posterior mean/s.d. ratio
Intercept	**0.076**	0.017	4.385	**0.038**	0.017	2.234
OECD				**0.036**	0.018	2.014
GDP60	**−0.018**	0.002	8.122			
Non-OECD				**−0.013**	0.002	5.717
OECD				**−0.013**	0.002	5.483
LifeExp60	**0.001**	0.000	4.829			
Non-OECD				**0.001**	0.000	7.616
OECD				**0.001**	0.000	7.094
EQINV	**0.148**	0.036	4.145			
Non-OECD				**0.156**	0.033	4.786
OECD				**0.156**	0.033	4.752
Mining	**0.033**	0.012	2.823			
Non-OECD				**0.046**	0.010	4.44
OECD				**0.046**	0.011	4.357
OutOrient	**−0.003**	0.002	1.644			
Non-OECD				**−0.003**	0.002	1.358
OECD				**−0.003**	0.002	1.38
LatAmDum	**−0.013**	0.005	2.756			
Non-OECD				**−0.016**	0.003	4.539
HighEd60	**−0.121**	0.029	4.093			
Non-OECD				**−0.192**	0.044	4.375
OECD				−0.012	0.029	0.424
SubSahAfricaDum	**−0.022**	0.004	5.143			
Non-OECD				**−0.014**	0.003	4.073
EthnoFrac	**0.015**	0.004	3.775			
Non-OECD				**0.020**	0.005	3.64
OECD				0.006	0.006	0.897
HinduFrac	**−0.108**	0.020	5.349			
Non-OECD				−0.016	0.019	0.856
Lforce60	**0.000**	0.000	4.924			
Non-OECD				0.000	0.000	0.607
SpainDum	**0.014**	0.005	2.799			
Non-OECD				NA	NA	NA
FrenchDum	**0.011**	0.004	2.71			
Non-OECD				0.002	0.003	0.882
NonEqInv	**0.031**	0.021	1.474			
Non-OECD				0.012	0.016	0.753
OECD				0.011	0.017	0.665
ConfuciousFrac	**0.074**	0.010	7.225			
Non-OECD				NA	NA	NA
EngLangFrac	**−0.007**	0.004	1.507			
Non-OECD				0.000	0.001	0.144
PrimaryEd60	**0.020**	0.009	2.268			
Non-OECD				NA	NA	NA
Civlibb	**−0.002**	0.001	2.038			
Non-OECD				NA	NA	NA
BritDum	**0.007**	0.003	2.394			
Non-OECD				NA	NA	NA

Table 1 (*continued*)

	Global			Interaction		
	Posterior mean	Posterior s.d.	Posterior mean/s.d. ratio	Posterior mean	Posterior s.d.	Posterior mean/s.d. ratio
RuleLaw	**0.013**	0.004	3.35			
Non-OECD				0.002	0.004	0.47
OECD				**−0.016**	0.011	1.453
BlackMktPrem	−0.004	0.004	1.216			
Non-OECD				**−0.012**	0.002	5.042
OECD				**−0.012**	0.002	5
EconOrg	0.000	0.001	0.567			
Non-OECD				**0.003**	0.001	4.235
OECD				**0.003**	0.001	3.927
PrimExp70	0.000	0.001	0.097			
Non-OECD				**−0.020**	0.004	5.014
OECD				**−0.017**	0.006	2.93
CathFrac	0.000	0.001	0.208			
Non-OECD				**0.013**	0.004	3.295
OECD				**0.013**	0.004	3.296
AvgPopAge	0.000	0.000	0.208			
Non-OECD				**0.000**	0.000	2.982
OECD				0.000	0.000	0.067
ProtFrac	−0.001	0.003	0.224			
Non-OECD				**−0.021**	0.010	2.083
OECD				0.004	0.005	0.723
BuddhaFrac	0.003	0.004	0.611			
Non-OECD				**0.018**	0.005	3.968
OthFracLang	0.000	0.001	0.108			
Non-OECD				**0.013**	0.003	3.99
OECD				−0.005	0.004	1.193

comparable to the specification used to establish the benchmark results for the global sample. In Section 4 we report robustness results that vary z.

The dependent variable is growth 1960–1992 and the first column of Table 1 features all regressors that were found to be effective (post. mean/s.d. > 1.3) in the global, OECD, or the Non-OECD samples.[15] Columns 2 and 3 report the coefficients for the global sample. For this sample no interaction terms are employed, hence the number of regressors is only 41, which allows the use of standard BMA algorithms. Of the 41 regressors considered, Table 1 reports only the relevant 31 regressors with post. mean/s.d. > 1.3 to save space. All regressors excluded from the tables are ineffective in the global sample, in all subsample analyses, and in all robustness specifications.

In the case of the global sample (columns 2, 3) no interaction terms are included, which implicitly assumes the absence of parameter heterogeneity. Here we replicate the results of the previous literature that assumes that OECD and Non-OECD countries are considered to have identical determinants of their growth performance, and that the magnitude of these determinants is also unchanged across subsamples. We find that in the global sample, 20 of

---

[15] Posterior coefficient estimates in bold font represent those variables that pass the effectiveness threshold (post. mean/s.d. > 1.3).

the 41 candidate variables are effective to growth. The number and the type of regressors that we identify as effective is in line with the findings of the previous literature. For example, Equipment Investment, Dummies relating to the colonial history, Initial GDP, and specific country characteristics matter to growth as in Sala-i-Martin et al. (2004); and FLS.

In columns 4 and 5 of Table 1 we report the results generated by allowing for the possibility of parameter heterogeneity related to the OECD group of countries. The subsample results are classified into seven subsets. First we have 5 variables that are effective in the global sample and in both the OECD and Non-OECD countries. These variables are Initial GDP, Initial Life Expectancy, Equity Investment, Mining and Outward orientation. This is the extent to which global, OECD and Non-OECD results agree. Second we find a set of 4 variables that are effective in both the global and Non-OECD samples, but are ineffective in the OECD sample. Variables in this set are Initial Higher Education, Ethnolinguistic Fragmentation, Sub-Saharan Africa, and the Latin Dummy. None of these variables have an impact in OECD countries. Two of these variables, the Sub-Saharan and Latin American Dummy, are simply irrelevant for OECD countries. For the other two the marginal contribution, β_2, in the interaction regression is highly significant and of the opposite sign as β_1, which renders the composite coefficient that indicates the OECD effect, $\bar{\beta}$, ineffective.

The third subset of results summarized in columns 4 and 5 of Table 1 is a relatively large set of 10 variables that are highly effective in the global sample, but once we allow for parameter heterogeneity neither the OECD nor the Non-OECD samples can claim these variables as growth determinants. Indeed in the interaction IBMA runs several of these variables do not pass the 1% posterior probability threshold and are not even included in the final regressor window that identifies the 41 top regressors. These cases are indicated with ''NA.''

The fourth category consists of only one variable, Rule of Law, which is effective in the global and OECD samples but ineffective in the Non-OECD sample. The fifth category consists of 4 variables that are not effective in the global sample but highly effective in both the OECD and Non-OECD subsamples. The Fraction of Catholics and the Degree of Capitalism (EconOrg) both have a positive effect in the OECD and Non-OECD sample while the Black Market Premium and Primary Exports have a negative effect on growth in OECD and Non-OECD countries.

The sixth category consists of 4 variables that are ineffective in the global sample, but effective only in Non-OECD countries. This result confirms that adding high-income countries to the global mix may drown out important effects in the developing country's subsample. The Average Population Age, the Fraction Protestant, Buddha and the Fraction of the Foreign Speaking Population are highly effective in Non-OECD countries but not in the global or OECD samples. Parameter heterogeneity thus uncovers not only crucial information as to what are not important growth determinants in advanced countries, but also new and important growth determinants in Non-OECD countries. Note that three of the variables that share importance in Non-OECD countries indicate a higher coefficient for the Non-OECD sample compared to the global sample. For two of these variables, Fraction Protestant and the Fraction of the Foreign Speaking Population, the impact in OECD countries is even opposite albeit ineffective. This is additional evidence that the inclusion of OECD countries in the sample drives down the growth impact of a variable for developing countries and may render it ineffective in the global sample. The seventh category consists of all variables that are ineffective in either the global, OECD or Non-OECD countries.

4. Robustness

The key innovation of IBMA is to apply the existing BMA structure iteratively to a computationally feasible subset of models, which we call the regressor window, z. In this section we examine the sensitivity of this novel aspect of IBMA analysis, as we vary the size of the regressor window. As indicated above, the previous growth literature established that between 4 and at most 18 variables matter in growth regressions, hence it would be surprising to obtain evidence from different window sizes that contradict our previous results. However, larger window sizes allow for more possible combinations of variables, some of which may not be able to attain the explanatory power unless they are placed in the models with a large number of regressors, yet others might not attain our threshold level of effectiveness unless they are jointly paired. The importance of such jointness has been emphasized by Doppelhofer and Weeks (2005) and Ley and Steel (in this issue). Table 2 reports the results for the global and the interacted sample from successively increasing regressor the window size. The practical computational limit is reached at a window size of $z = 48$.

To present the results most efficiently we have combined two columns in Table 1 to one individual column per window size that reports the global, OECD, and Non-OECD estimates for each relevant variable. Note that the Global estimate is only provided as a reference; it does not change throughout since the models for all 41 variables can be examined in BMA. Only the interaction that separates OECD and Non-OECD increases the number of regressors from 41 to 77, requiring the application of IBMA. Overall Table 2 documents robust results, but there are important changes that we discuss in detail.

Moving from $z = 41$ to $z = 45$ generates only a few differences in the results. For OECD countries we now find Mining to be ineffective while Non-Equipment Investment becomes effective. Additionally, we now find the Average Population Age, Fraction Protestant and the share of the Workforce to Total Population to be highly effective for OECD countries. For Non-OECD countries there are only two changes among the 41 growth determinants. The two additional variables that now register as marginally effective for Non-OECD Countries are Non-Equipment Investment and the Fraction Hindu, but otherwise there is no difference in the results. Most convincing perhaps is that the coefficient estimates are just about unchanged.

As we increased the size of the regressors window past $z = 45$, we find slightly augmented results. For the computationally most demanding run, $z = 48$, we find that a greater number of variables matter in both the global and the Non-OECD sample. Allowing for a larger window size increased the explanatory power for the Non-OECD determinants initial Labor Force, the Hindu Dummy, the Spanish Dummy, the French Dummy, and Non-Equipment Investment; every one of these variables was initially effective in the global sample, but ineffective in either subsample. In addition, the War Dummy and Property Rights are now also effective for Non-OECD, although they are not effective for the global sample. Two variables, the Fraction Catholic and Protestant now become ineffective. For OECD countries there are also a number of changes as 8 additional variables are added to the list of effective variables while 3 (Mining, Rule of Law and Fraction Catholic) are dropped from this list. On balance, however, the picture is unchanged as the evidence for parameter heterogeneity is overwhelming.

The clear break that signifies a large increase in the variables that are effective is at $z = 45$. After $z = 45$ (see example for $z = 47$) the results are all closer to $z = 48$ than to

$z = 41$, that a effectively larger number of variables matters for growth in OECD and Non-OECD countries. However, we cannot identify a single variable that remains uniquely effective for OECD countries across the different window sizes. This is perhaps yet again more evidence that this dataset does not contain variables that are the unique growth determinants in this subset of countries.

The conclusions that can be reached from our robustness exercise are twofold. First, most of our important benchmark results are quite robust to changes in the size of the regressor window. We caution though that these results have also revealed some fragility inherent in the regressor window approach inherent in the IBMA methodology. This should be kept in mind when one assigns particular interpretation to certain variables. Scrutinizing the causes for possible fragility of IBMA is beyond of the scope of this paper but we judge this as an important area for future research.

5. Discussion

In general our results suggest that the important determinants of long-term growth in Non-OECD countries overlap only to some degree with the factors identified with the global samples. For OECD countries this overlap is even smaller. In addition, allowing for parameter heterogeneity unveiled a large number of new variables that matter to only Non-OECD countries. However, allowing for parameter heterogeneity did not allow us to gain any meaningful insights into unique factors that determine growth in OECD countries.

We provide a Summary Table 1a to collect the results. Overall we find that a number of purported growth determinants in the global sample are not effective for Non-OECD countries, and that most established growth determinants do not show explanatory power for OECD countries. Even for Non-OECD countries, 11 of the original 20 effective variables are no longer effective. Instead, an entirely new set of variables matters in Non-OECD Countries, where 8 variables that were ineffective in the global sample are now shown to matter. While it is surprising to see some of the key variables in the global sample, such as Civil Liberties, Fraction Confucius, and Primary Education, loose their significance, the newly effective variables are all very much in line with established key indicators of growth developing nations, such as the Degree of Capitalism, Primary Exports Share, and the Black Market Premium.

For the OECD the results are even more stunning. Of all the original 20 effective variables in the global sample only 6 survive as effective. The only variables added as effective for OECD countries by allowing for parameter heterogeneity are the Fraction of Population that is Catholic, Primary Exports, the Degree of Capitalism, and the Black Market Premium. The evidence for parameter heterogeneity is therefore overwhelming. Most variables in the global dataset do not matter for OECD countries, and half of the variables that matter for Non-OECD countries also do not matter for OECD countries. Note that this implies (as per our discussion in Section 3.2) that the OECD treatment effect is highly significant *and of the opposite sign* as the Non-OECD effect to render the composite coefficient for the OECD, β, insignificant.

The combined analysis of parameter heterogeneity and model uncertainty has lead not only to quantitative differences regarding the effect of growth determinants across subsamples, but it also generated important new qualitative implications. To our surprise the quantitative (economic) differences between subsamples were minimal, because so few regressors

Table 1a
Summary of effective growth determinants

	BMA		IBMA with interactions			
	$y = \alpha \iota_n + Z_j \beta_j + \sigma \varepsilon$		$y = \alpha \iota_n + Z_j \beta_{1,j} + I\, X_j \beta_{2,j} + \sigma \varepsilon$			
	Global sample effective variables		Non-OECD effective variables		OECD effective variables	
	Posterior mean	Posterior s.d.	Posterior mean	Posterior s.d.	Posterior mean	Posterior s.d.
BritDum	0.007	0.003				
Civlibb	−0.002	0.001				
ConfuciousFrac	0.074	0.01				
EngLangFrac	−0.007	0.004				
PrimaryEd60	0.02	0.009				
NonEqInv	0.031	0.021				
FrenchDum	0.011	0.004				
Lforce60	0.000	0.000				
HinduFrac	−0.108	0.02				
SpainDum	0.014	0.005				
LatAmDum	−0.013	0.005	−0.016	0.003		
FracEthno	0.015	0.004	0.02	0.005		
SubSahAfricaDum	−0.022	0.004	−0.014	0.003		
HighEd60	−0.121	0.029	−0.192	0.044		
EQInvest	0.148	0.036	0.156	0.033	0.156	0.033
LifeExp60	0.001	0.000	0.001	0.000	0.001	0
OutOrient	−0.003	0.002	−0.003	0.002	−0.003	0.002
Mining	0.033	0.012	0.046	0.01	0.046	0.011
GDP60	−0.018	0.002	−0.013	0.002	−0.013	0.002
RuleLaw	0.013	0.004			−0.016	0.011
CathDum			0.013	0.004	0.013	0.004
PrimExp70			−0.02	0.004	−0.017	0.006
BlackMktPrem			−0.012	0.002	−0.012	0.002
EconOrg			0.003	0.001	0.003	0.001
BuddhaDum			0.018	0.005		
AvgPopAge			0.000	0.000		
OthFracLang			0.013	0.003		
ProtFrac			−0.021	0.01		

Note: Column 4 reports composite means and the associated composite standards deviations. All variables that do not meet our effectiveness threshold (post. mean/s.d. < 1.3) are not reported to save space.

are common across subsamples. Qualitatively we find not only that regressors may have opposite impacts in the different subsamples, but that indeed an entirely different set of regressors matters in the global, Non-OECD and OECD samples. While the relevant regressors for the global and Non-OECD sample can be recovered, the dataset does not contain the regressors necessary to explain the OECD growth performance. This is doubly tragic. First, policy recommendations to lower income countries can no longer be framed within the context that improvements in any of the variables in the dataset will actually lead to better growth outcomes. Hence we have no guidance as to what drives growth in high income countries. But even more disturbing, the growth performance in OECD countries was on average twice as high as in the Non-OECD samples, hence neither the determinants of the higher income levels, nor the higher income growth rates can be recovered given the current dataset. Two avenues can be explored to reconcile these findings. First we can collect data that has been linked

Table 2
Robustness using different window sizes in IBMA

	Regressor window size 41		Regressor window size 45		Regressor window size 47		Regressor window size 48	
	Posterior mean	Posterior s.d.	Posterior Mean	Posterior s.d.	Posterior mean	Posterior s.d.	Posterior mean	Posterior s.d.
Intercept	**0.038**	0.017	**0.038**	0.017	**0.071**	0.014	**0.070**	0.015
OECD	**0.036**	0.018	**0.035**	0.021	**0.077**	0.026	**0.059**	0.027
GDP60	**−0.018**	0.002	**−0.018**	0.002	**−0.018**	0.002	**−0.018**	0.002
Non-OECD	**−0.013**	0.002	**−0.012**	0.002	**−0.015**	0.002	**−0.015**	0.002
OECD	**−0.013**	0.002	**−0.013**	0.002	**−0.015**	0.002	**−0.014**	0.002
LifeExp60	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000
Non-OECD	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000
OECD	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000	**0.001**	0.000
EQINV	**0.148**	0.036	**0.148**	0.036	**0.148**	0.036	**0.148**	0.036
Non-OECD	**0.156**	0.033	**0.156**	0.033	**0.179**	0.032	**0.181**	0.034
OECD	**0.156**	0.033	**0.156**	0.033	0.052	0.046	**0.091**	0.063
Mining	**0.033**	0.012	**0.033**	0.012	**0.033**	0.012	**0.033**	0.012
Non-OECD	**0.046**	0.010	**0.041**	0.011	**0.029**	0.009	**0.027**	0.010
OECD	**0.046**	0.011	0.036	0.028	**0.032**	0.020	−0.016	0.067
OutOrient	**−0.003**	0.002	**−0.003**	0.002	**−0.003**	0.002	**−0.003**	0.002
Non-OECD	**−0.003**	0.002	**−0.003**	0.002	**−0.002**	0.001	**−0.003**	0.001
OECD	**−0.003**	0.002	**−0.003**	0.002	−0.002	0.002	**−0.003**	0.002
LatAmDum	**−0.013**	0.005	**−0.013**	0.005	**−0.013**	0.005	**−0.013**	0.005
Non-OECD	**−0.016**	0.003	**−0.016**	0.004	**−0.011**	0.003	**−0.013**	0.003
HighEd60	**−0.121**	0.029	**−0.121**	0.029	**−0.121**	0.029	**−0.121**	0.029
Non-OECD	**−0.192**	0.044	**−0.200**	0.047	**−0.111**	0.032	**−0.120**	0.029
OECD	−0.012	0.029	−0.025	0.033	**0.038**	0.028	−0.119	0.030
SubSahAfrica	**−0.022**	0.004	**−0.022**	0.004	**−0.022**	0.004	**−0.022**	0.004
Non-OECD	**−0.014**	0.003	**−0.016**	0.004	**−0.016**	0.002	**−0.017**	0.003
EthnoFrac	**0.015**	0.004	**0.015**	0.004	**0.015**	0.004	**0.015**	0.004
Non-OECD	**0.02**	0.005	**0.021**	0.005	**0.017**	0.004	**0.017**	0.004
OECD	0.006	0.006	−0.002	0.006	−0.003	0.005	0.000	0.006
HinduFrac	**−0.108**	0.020	**−0.108**	0.020	**−0.108**	0.020	**−0.108**	0.020
Non-OECD	−0.016	0.019	**−0.031**	0.020	**−0.083**	0.019	**−0.068**	0.013
Lforce60	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
Non-OECD	0.000	0.000	0.000	0.000	**0.000**	0.000	**0.000**	0.000
OECD	NA	NA	NA	NA	**0.000**	0.000	**0.000**	0.000
SpainDum	**0.014**	0.005	**0.014**	0.005	**0.014**	0.005	**0.014**	0.005
Non-OECD	NA	NA	NA	NA	**0.010**	0.003	**0.009**	0.003
FrenchDum	**0.011**	0.004	**0.011**	0.004	**0.011**	0.004	**0.011**	0.004
Non-OECD	0.002	0.003	0.002	0.002	**0.007**	0.003	**0.005**	0.002
NonEqInv	**0.031**	0.021	**0.031**	0.021	**0.031**	0.021	**0.031**	0.021
Non-OECD	0.012	0.016	**0.027**	0.017	**0.054**	0.014	**0.054**	0.014
OECD	0.011	0.017	**0.027**	0.017	−0.016	0.025	**−0.050**	0.029
EngLangFrac	**−0.007**	0.004	**−0.007**	0.004	**−0.007**	0.004	**−0.007**	0.004
Non-OECD	0.000	0.001	NA	NA	**−0.018**	0.006	−0.010	0.004
OECD	NA	NA	NA	NA	−0.001	0.004	NA	NA
Civlibb	**−0.002**	0.001	**−0.002**	0.001	**−0.002**	0.001	**−0.002**	0.001
Non-OECD	NA	NA	NA	NA	0.000	0.001	NA	NA
OECD	NA	NA	NA	NA	**0.008**	0.004	NA	NA
BritDum	**0.007**	0.003	**0.007**	0.003	**0.007**	0.003	**0.007**	0.003
Non-OECD	NA	NA	NA	NA	0.003	0.003	0.000	0.001
OECD	NA	NA	NA	NA	NA	NA	**0.017**	0.004

Table 2 (*continued*)

	Regressor window size 41		Regressor window size 45		Regressor window size 47		Regressor window size 48	
	Posterior mean	Posterior s.d.	Posterior Mean	Posterior s.d.	Posterior mean	Posterior s.d.	Posterior mean	Posterior s.d.
RuleLaw	**0.013**	0.004	**0.013**	0.004	**0.013**	0.004	**0.013**	0.004
Non-OECD	0.002	0.004	0.006	0.005	NA	NA	NA	NA
OECD	**−0.016**	0.011	**−0.037**	0.013	NA	NA	NA	NA
BlkMktPrem	−0.004	0.004	−0.004	0.004	−0.004	0.004	−0.004	0.004
Non-OECD	**−0.012**	0.002	**−0.011**	0.002	**−0.005**	0.002	**−0.006**	0.002
OECD	**−0.012**	0.002	**0.027**	0.008	0.005	0.011	**0.027**	0.007
EconOrg	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	**0.003**	0.001	**0.003**	0.001	**0.002**	0.001	**0.002**	0.000
OECD	**0.003**	0.001	**0.003**	0.001	**0.002**	0.001	**0.002**	0.001
PrimExp70	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	**−0.02**	0.004	**−0.021**	0.004	**−0.020**	0.003	**−0.021**	0.003
OECD	**−0.017**	0.006	**−0.021**	0.004	**−0.020**	0.003	**−0.021**	0.003
CathFrac	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	**0.013**	0.004	**0.013**	0.004	NA	NA	NA	NA
OECD	**0.013**	0.004	**0.013**	0.004	NA	NA	NA	NA
AvgPopAge	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Non-OECD	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
OECD	0.000	0.000	**0.000**	0.000	0.000	0.000	**0.000**	0.000
ProtFrac	−0.001	0.003	−0.001	0.003	−0.001	0.003	−0.001	0.003
Non-OECD	**−0.021**	0.010	**−0.021**	0.008	−0.003	0.004	−0.001	0.004
OECD	0.004	0.005	**0.009**	0.005	−0.004	0.004	0.001	0.003
BuddhaFrac	0.003	0.004	0.003	0.004	0.003	0.004	0.003	0.004
Non-OECD	**0.018**	0.005	**0.018**	0.005	**0.019**	0.004	**0.015**	0.004
OthFracLang	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	**0.013**	0.003	**0.012**	0.003	**0.010**	0.002	**0.010**	0.003
OECD	−0.005	0.004	0.000	0.004	−0.004	0.004	**−0.005**	0.003
PropRights	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	NA	NA	NA	NA	**−0.002**	0.001	**−0.002**	0.001
OECD	NA	NA	NA	NA	−0.003	0.004	**−0.002**	0.001
WarDum	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001
Non-OECD	NA	NA	NA	NA	**−0.004**	0.001	−0.003	0.002
OECD	NA	NA	NA	NA	−0.003	0.002	0.000	0.005
Worker/Pop	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Non-OECD	NA	NA	0.000	0.001	NA	NA	0.000	0.000
OECD	NA	NA	**−0.036**	0.009	NA	NA	**−0.037**	0.014

Note: Posterior coefficient estimates in bold font represent variables that pass our effectiveness threshold (post. mean/s.d. > 1.3).

specifically to growth in OECD countries (for example on regulation, see Nicoletti and Stephan Scarpetta, 2003). However, hopes of expanding such a dataset to the global sample are perhaps unrealistic. Second, the notion of one size fits all – or that one theory or one approach to growth can address the growth determinants in disparate subsamples – might be too optimistic.

6. Conclusion

This paper extends the literature on country heterogeneity in two dimensions. First, a new model averaging method called Iterative Bayesian Model Averaging (IBMA) is used

to handle the exhaustive computation required when we simultaneously consider model uncertainty and parameter heterogeneity in our estimation. Second, instead of investigating the sources of growth (or lack of it) in low-income countries, we take a fresh look at what determines growth performance in the high-income OECD countries.

Our analysis suggests that IBMA is a powerful technique that makes it possible for researchers to consider a very large number of potential regressors. Our application of IBMA to growth empirics allows us to examine parameter heterogeneity and model uncertainty simultaneously in all regressor candidates. It reveals that a large number of regressors is highly effective for Non-OECD countries, but irrelevant for both, OECD countries and the global sample. Perhaps most surprising was our finding that the long list of growth determinants included in popular cross-country datasets does not contain variables that begin to identify the key determinants of growth in advanced countries. "Global" results that have been taken to represent some average coefficient estimate for all countries are now shown to provide little information about the growth determinants in two key subsamples. The Global results have been debunked as artefacts of the combination of two heterogeneous subsamples, and no longer as an expected impact that can identify effective growth determinants.

Acknowledgement

Appendix

See Table A1.

Table A1
Descriptive statistics

Variable	Global		OECD		Non-OECD	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Absolute latitude	25.733	17.250	45.126	10.461	16.630	11.189
Age	23.708	37.307	39.043	41.877	16.510	33.006
Area (scale effect)	972.917	2051.976	1467.130	3036.055	740.939	1353.317
Black market premium	0.157	0.291	0.059	0.196	0.203	0.318
British colony	0.319	0.470	0.174	0.388	0.388	0.492
Civil Liberties	3.466	1.712	1.758	1.148	4.268	1.295
Equipment invest.	0.044	0.035	0.072	0.024	0.031	0.031
Ethnolinguistic fractionalization	0.371	0.296	0.217	0.211	0.443	0.304
Fraction Catholic	0.422	0.397	0.427	0.392	0.420	0.403
Fraction of Buddhist	0.056	0.184	0.045	0.183	0.061	0.186
Fraction of Confucian	0.019	0.087	0.026	0.125	0.016	0.064
Fraction of Foreign speaking pop.	0.374	0.422	0.308	0.420	0.406	0.424
Fraction of Hindu	0.018	0.101	0.000	0.000	0.027	0.122
Fraction of Jews	0.013	0.097	0.002	0.005	0.018	0.117
Fraction of mining to GDP	0.045	0.077	0.017	0.018	0.058	0.090
Fraction of Muslim	0.148	0.295	0.044	0.208	0.196	0.318
Fraction of pop. speaking English	0.076	0.239	0.181	0.357	0.026	0.136
Fraction of Protestants	0.173	0.252	0.323	0.357	0.103	0.139

Table A1 (*continued*)

Variable	Global		OECD		Non-OECD	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Fraction of years open	0.439	0.355	0.737	0.203	0.299	0.325
French colony	0.125	0.333	0.000	0.000	0.184	0.391
GDP per capita 1960 (log)	7.492	0.885	8.399	0.622	7.066	0.633
Growth rate of population	0.020	0.010	0.009	0.007	0.026	0.006
Higher education enrolment, 1960	0.043	0.052	0.087	0.061	0.023	0.030
Latin American Dummy	0.278	0.451	0.043	0.209	0.388	0.492
Life expectancy, 1960	56.581	11.448	67.948	5.986	51.245	9.298
Non-equipment invest.	0.149	0.055	0.183	0.037	0.134	0.055
Outward orientation	0.389	0.491	0.435	0.507	0.367	0.487
Per Capita GDP growth 1960–1992	0.021	0.018	0.030	0.011	0.017	0.019
Political rights	3.451	1.896	1.589	0.993	4.324	1.558
Pop. 60* Worker 60 (Scale Effect)	9305.375	24906.050	12814.540	16980.030	7658.217	27869.810
Primary exports, 1970	0.673	0.299	0.379	0.230	0.811	0.217
Primary school enrolment, 1960	0.795	0.246	0.971	0.066	0.713	0.256
Public education share	0.025	0.009	0.029	0.010	0.022	0.008
Ratio of worker to pop (log)	−0.954	0.189	−0.885	0.132	−0.986	0.204
Real exchange rate distortion	121.708	41.001	105.783	16.605	129.184	46.709
Revolutions and coups	0.182	0.238	0.071	0.122	0.235	0.261
Rule of law	0.551	0.335	0.899	0.179	0.388	0.258
Spanish colony	0.222	0.419	0.043	0.209	0.306	0.466
Standard Deviation of BMP	45.596	95.802	3.190	7.512	65.500	110.832
Sub-Saharan African Dummy	0.208	0.409	0.000	0.000	0.306	0.466
Type of econ. organization	3.542	1.266	4.217	0.736	3.224	1.343
War Dummy	0.403	0.494	0.130	0.344	0.531	0.504
Number of obs.	**72**		**23**		**49**	

Note: For Civil Liberties and Political Rights higher values imply lower civil liberties and political rights.

References

Azariadis, C., Drazen, A., 1990. Threshold externalities in economic development. Quarterly Journal of Economics, 501–526.

Brock, W., Durlauf, S., 2001. Growth empirics and reality. The World Bank Economic Review 15, 229–272.

Brock, W., Durlauf, S., West, K., 2003. Policy evaluation in uncertain economic environments. Brooking Papers on Economic Activity 1, 235–322.

Doppelhofer, G, Weeks, M, 2005. "Jointness of Determinants of Economic Growth," Working Paper, Cambridge University.

Durlauf, S.N., Johnson, P., Temple, J., 2005. Growth econometrics. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth. Amsterdam, North Holland.

Draper, P., 1995. Assessment and propagation of model uncertainty, (with discussion). Journal of the Royal Statistical Society B 57, 45–97.

Easterly, W., Levine, R., 1997. Africa's growth tragedy: Policies and ethnic divisions. Quarterly Journal of Economics 112, 1203–1250.

Eicher, T.S., Leukert, A., 2006. "Institutions and Economic Performance: Endogeneity and Parameter Heterogeneity," Munich discussion paper 2006-05.

Eicher, T.S., Papageorgiou, C., Raftery, A.E., 2007, "Evaluating the Impact of Priors in Economic Applications of Bayesian Model Averaging," working paper, University of Washington.

Fernandez, C., Ley, E., Steel, M., 2001a. Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics 16, 563–576.

Fernandez, C., Ley, E., Steel, M., 2001b. Benchmark priors for bayesian model averaging. Journal of Econometrics 100, 381–427.

Furnival, G.M., Wilson, R.W., 1974. Regression by leaps and bounds. Techonometrics 16, 499–511.

Galor, O., Moav, O., 2002. Natural selection and the origin of economic growth. Quarterly Journal of Economics 117, 1133–1192.

Galor, O., Weil, D.N., 2000. Population, technology, and growth: From the malthusian regime to the demographic transition and beyond. American Economic Review 90, 806–828.

Hoeting, J., Madigan, D., Raftery, A.E., Volinsky, C., 1999. "Bayesian Model Averaging: A Tutorial," (with discussion). Statistical Science, 14, 382–401.

Kass, R.E., Wasserman, L., 1995. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. Journal of the American Statistical Association 90, 928–934.

Leon-Gonzalez, R., Montolio, D., 2004. Growth, convergence and public investment. A bayesian model averaging approach. Applied Economics 36, 1925–1936.

Leamer, E.E., 1978. Specification Searches. Wiley, New York, NY.

Leamer, E.E., 1983. Let's take the con out of econometrics. American Economic Review 73, 31–43.

Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. American Economic Review 82, 942–963.

Ley, E., Steel M., this issue, "Jointness in variable selection with applications to growth regression," Journal of Macroeconomics.

Lucas, R.E., 2002. Lectures on Economic Growth. Harvard University Press, Cambridge.

Madigan, D., York, J., 1995. Bayesian graphical models for discrete data. International Statistical Review 63, 215–232.

Masanjala, W.H., Papageorgiou, C., 2007a, "A Rough and Lonely Road to Prosperity: A Reexamination of Sources of Growth in Africa using Bayesian Model Averaging," working paper, LSU.

Masanjala, W.H. and C. Papageorgiou, 2007b, "Initial Conditions and Post-War Growth in sub-Saharan Africa," working paper, LSU.

Mitchell, T.J., Beauchamp, J.J., 1988. Bayesian variable selection in linear regression. Journal of the American Statistical Association 83, 1023–1036.

Multon, B.R., 1991. A bayesian approach to regression selection and estimation with application to a price index for radio services. Journal of Econometrics 49, 169–193.

Nicoletti, G., Stephan Scarpetta, S., 2003. Regulation, productivity and growth: OECD evidence. Economic Policy 18 (1), 9–72 (64).

Palm, F.C., Zellner, A., 1992. To combine or not to combinena issues of combining forecasts. Journal of Forecasting 11, 687–701.

Raftery, A.E., 1995. Bayesian model selection for social research. Sociological Methodology 25, 111–163.

Raftery, A.E., 1999. Bayes factors and BIC: Comment on weakliem. Sociological Methods and Research 27, 411–427.

Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian model averaging for linear regression models. Journal of the American Statistical Association 92, 179–191.

Sala-i-Martin, X.G., 1997. I just ran two million regressions. American Economic Review, Papers and Proceedings 87, 178–183.

Sala-i-Martin, X.G., Doppelhofer, G., Miller, R., 2004. Determinants of long-term growth: A bayesian averaging of classical estimates (BACE) approach. American Economic Review 94, 813–835.

Temple, J., 2000. Growth regressions and what the textbooks do not tell you. Bulletin of Economic Research 52, 181–205.

Volinsky, C.T., Madigan, D., Raftery, A.E., Kronmal, R.A., 1997. Bayesian model averaging in proportional hazard models: Assessing the risk of a stroke. Journal of Royal Statistical Society 46, 433–448.

Yeung, K.Y., Baumgarner, R.E., Raftery, A.E., 2005. Bayesian model averaging: Development of an improved multi-class, gene selection and classification tool for microarray data. Bioinformatics 21, 2394–2402.