CENTRO DE ESTUDIOS MONETARIOS Y FINANCIEROS (CEMFI)
UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO (UIMP)

ESSAYS ON GROWTH ECONOMETRICS

by

ENRIQUE MORAL-BENITO

DOCTORAL DISSERTATION

June 2010

CENTRO DE ESTUDIOS MONETARIOS Y FINANCIEROS (CEMFI)
UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO (UIMP)

ESSAYS ON GROWTH ECONOMETRICS

by

ENRIQUE MORAL-BENITO

Supervisor:

Manuel Arellano

DOCTORAL DISSERTATION

June 2010

A mis padres

Acknowledgements

This is perhaps the easiest and hardest chapter that I have to write. It is simple to name all the people that helped to get this dissertation done, but it is almost impossible to thank them enough.

First, it is difficult to overstate my gratitude for my PhD supervisor, Manuel Arellano. With his efforts to explain things clearly and simply, his great patience answering my questions, and his infinite knowledge of economics and econometrics, he helped me to understand what I was doing. Throughout the elaboration of this thesis, he provided invaluable advice, good teaching, and lots of good ideas. I would have been lost without him, and it has been a great honor to work with him.

Next, I would like to thank the other members of my committee, Stéphane Bonhomme, Enrique Sentana, and Claudio Michelacci, for many enlighting conversations and friendly support, and the rest of faculty at CEMFI for always holding the door open. I am also grateful to Rafael Repullo and Antonio Álvarez Pinilla for provoking my interest in working towards this PhD dissertation. During my days at University of Oviedo, Antonio was the first in seeing me as a potential researcher, I will always be indebted to him. Likewise, I am also grateful to the Infraestructure and Transport Research Group at University of Las Palmas. More concretely, special thanks go to Ginés de Rus, who substantially contributed to my training as economist.

Furthermore, I would like to express my gratitude to Luis Servén and Eduardo Ley for their hospitality and support during my visits to The World Bank. With Luis I learned a lot about macroeconomics and panel cointegration, and Eduardo decisively enhanced my understanding of Bayesian econometrics. I am also thankful for their comments and suggestions on different drafts of the chapters of this dissertation.

I am indebted to my PhD colleagues at CEMFI for providing a pleasant environment in which to learn and grow over the last three years. Alicia Barroso, Enzo Cerletti, Manuel García, Ainara González, Pablo Lavado, Cristina López-Mayán, David Martínez-Miera, Jorge de la Roca, and Hernán Ruffo positively contributed to make these years more enjoyable. Cristian Bartolucci, Carlos Gonález-Aguado, Joan Llull, and Roberto Ramos deserve special mention for their comments on several drafts and, more importantly, for so many hours of cheap talk and for pushing the frontier during the coffee breaks.

I am also grateful to the staffs at CEMFI, UIMP and The World Bank, for helping the formalities run smoothly and for assisting me in many different ways. I must give my special thanks to Ana Pérez, Araceli Requerey, Eloy Moros, Eva González, Gema Salazar, Irene Telo, Isabel Redondo, Javier Palomino, Mari Carmen Mondéjar, Miguel Benavente, Tourya Tourougui, and Ulises Tineo.

During my years as PhD student I have been fortunate enough to share my time off with many great people. Bel(l)man Friends are the best example in this category. I wish to thank the entire team for so many unforgettable Sundays in La Chopera. Eduard, Máximo, Nacho, Óscar, Pedro, and Roberto have all contributed to my formation as a friend.

Doy gracias a mi madre, Montse, por haber estado siempre ahí. Te ha tocado hacer de madre y padre a la vez, y lo has hecho a la perfección. Me has criado, me has apoyado, me has educado, me has querido, y sobre todo, me has enseñado a ser mejor persona. Gracias por todo Mamá. También quiero agradecer a mi hermano David por hacer que los momentos difíciles fueran menos difíciles. Estoy muy orgulloso de ti.

Lastly, and most importantly, I wish to thank my fiancée Marta for the very special person she is, and for the incredible amount of patience she has with me. It would take another thesis to express my deep love for you.

Contents

List of Tables

List of Figures

Introduction

For centuries economists have been preoccupied with the growth of nations. Although this effort has produced a better understanding of the sources of economic growth, the subject has proved elusive. In fact, since the late eighties, economists have shifted their focus from theory to estimation of linear growth regressions in an attempt to identify the key determinants of growth. This is not to say that growth theories are of no use for that purpose. Rather, the problem is that growth theories are, using a term due to Brock and Durlauf (2001), open-ended. This means 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. This model uncertainty (i.e. diversity of theoretical views) makes it hard to identify the most effective growth-promoting policies.

Related to this issue there is the convergence debate. The growth-regression apparatus has also been used to shed light on this debate. While some authors consider that the available empirical evidence supports the conditional convergence hypothesis predicted by the neoclassical growth model, skeptical researchers are more cautious given the absence of reliable estimates of the convergence parameter in growth regressions. All in all, after more than two decades of research the question is still unclear: are contemporary differences in growth rates across countries transient over sufficiently long time horizons? (i.e. is there conditional convergence across countries?).

Model uncertainty and endogeneity are responsible for the lack of consensus regarding these two issues in the empirical growth literature. Model uncertainty emerges because theory does not provide enough guidance to select the proper empirical model. Endogeneity arises because growth regressors are typically correlated with unobservables influencing economic growth. Although papers in the literature address both issues separately, there is no alternative proposed to address both at the same time. The main and overall objective of this doctoral thesis is to simultaneously overcome these two problems and thus shed light on the growth determinants and convergence debates. For this purpose, I introduce a novel panel maximum-likelihood estimator that overcomes the issue of endogeneity in the growth framework (Chapter 3), and I also show how to address model uncertainty from a panel-setting perspective taking into account endogeneity (Chapter 2).

Model uncertainty and its associated problems from an empirical viewpoint were first discussed by Leamer (1983). In particular, he pointed out the fragility of regression analysis to whimsical assumptions and arbitrary decisions about choice of control variables when model uncertainty is present (as it is clearly the case in growth empirics). His proposed remedy was sensitivity analysis of regression estimates that separates the fragile inferences from the solid ones. Model averaging techniques represent an alternative approach to address the issue of fragility in regression analysis. Imagine a situation in which there are many different candidate models for estimating the effect of X (e.g. investment) on Y (e.g. GDP growth). The most common approach is to estimate a single model, and then make inference based on that selected model ignoring the uncertainty surrounding the selection of that single model. A better alternative, however, would be to estimate all the candidate models, and then compute a weighted average of all the estimates for the coefficient on investment (X). Moreover, after computing the associated standard errors, one could make inference based on the whole universe of candidate models. By doing so, we would be considering not only the uncertainty associated to the parameter estimate conditional on a given model, but also the uncertainty of the parameter estimate across different models. This approach would lead us to more reliable, or at least more honest, conclusions regarding the significance of the estimated effect of investment on economic growth. Sala-i-Martin et al. (2004) and Fernández et al. (2001b) are good examples of the application of model averaging techniques to the growth regressions literature. Chapter 1 of this dissertation presents an overview of model averaging in economics.

In Chapter 2 I illustrate how to combine different panel data estimators with model averaging methods in order to simultaneously overcome model uncertainty and different forms of endogeneity. The main focus of Chapter 2 is the extension of model averaging to panel data settings with an application to growth regressions. Here, the regressors are assumed to be correlated with the individual and time-invariant effects but exogenous in the reverse causality sense. Therefore I refer to them as panel growth covariates rather than panel growth causes. I draw two main conclusions: On the one hand, the empirical findings suggest that country specific effects correlated with other regressors play an important role, because the list of robust growth covariates is not the same as when we do not take their presence into account. On the other hand, the empirical estimate of the rate of convergence, after controlling for both model uncertainty and endogeneity of the lagged dependent variable,[1] is surprisingly similar to that commonly found in cross-section studies but not significantly different from zero.

Endogeneity problems with estimating an empirical growth model are well known. The right-hand side variables are typically endogenous and measured with error. Omitted variable bias also arises because of the presence of unobservable time-invariant country-specific characteristics correlated with one or more regressors. The most prominent way to address these problems is the use of panel data econometric techniques that allow for country-specific fixed effects in the empirical model. In particular, first-differenced GMM estimators applied to dynamic panel data models has been the most promising econometric method in empirical growth research. This estimation procedure addresses the question of correlated individual effects and the issue of endogeneity. It was first proposed in the econometrics literature by Holtz-Eakin et al. (1988) and Arellano and Bond (1991). In the growth context, the method was first considered by Caselli et al. (1996). Despite its important advantages over simple cross-section regressions and other estimation methods for dynamic panel-data models, it is now well known that in the growth context this method suffers from large finite-sample biases because of the many weak instruments problem. To overcome this issue, I present in Chapter 3 a likelihood-based estimator in a panel data context that is asymptotically equivalent to one-step first-differenced GMM. This maximum likelihood estimator alleviates the weak-instruments problem in finite samples without resorting to auxiliary stationarity assumptions. Having endogenous regressors allows me to tackle the issue of reverse causality and therefore argue that the resulting estimates can be interpreted as causal effects. Moreover, since it is as general as first differenced GMM, the estimator can be applied to a broad range of situations in addition to growth regressions. Via Monte Carlo simulations I conclude that it is strongly preferred to standard panel GMM in terms of finite-sample performance.

Furthermore, in Chapter 3 I also combine this estimator with model averaging techniques

[1]Despite the regressors are exogenous in Chapter 2, the endogeneity of the lagged dependent variable generated by the dynamics of the model is taken into account.

employing the so-called Bayesian Averaging of Maximum Likelihood Estimates (BAMLE) in the framework of growth regressions. Considering these two ingredients, I argue that this is the first attempt in the literature to simultaneously consider endogenous regressors and model uncertainty. It is important to remark that despite the focus on growth empirics, the proposed methodology can be useful in other settings in which endogeneity and model uncertainty are causes of concern.

Given the above, after considering all potential sources of biases and inconsistencies (i.e. after combining the model averaging methodology with the proposed likelihood-based estimator), I obtain two results that are in contrast to the previous consensus in the empirical growth literature. On the one hand, I find that conditional convergence is not present across the countries in my sample. In particular, the estimated speed of convergence is 0.73%, but it is not significantly different from zero. This result would lead us to conclude that the hypothesis of no conditional convergence cannot be rejected given the available data. On the other hand, I conclude that there is only one variable that seems to have a causal connection to economic growth, the investment ratio.

CHAPTER 1

Model Averaging in Economics: A Survey

1.1 Introduction

In the early Eighties, Leamer (1983) pointed out the fragility of regression analysis to arbitrary decisions about choice of control variables. The selection of control variables in empirical research is an important issue which usually hampers consensus on the particular empirical model to be estimated. The empirical growth literature is probably the best example. In the growth regressions industry, 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. Parameter estimates emerging from these regressions are highly fragile to the inclusion of different sets of regressors.

Extreme Bound Analysis (EBA) was first proposed (Leamer (1983), Leamer and Leonard (1983)) as a tool for quantifying the sensitivity of regression estimates. Suppose one is interested in measuring the effect of the variable X on the variable Y. Extreme Bounds Analysis consist of the following steps: first, we estimate a fairly general model in which you regress Y on X and a set of other (control) variables; second, we estimate several simplified versions of the general model (for example by excluding one or more explanatory variables); finally we analyze all the different estimated coefficients on X. Extreme Bounds Analysis is concerned with the largest and smallest values of these estimates. Suppose the estimated coefficient varies greatly over the range of estimated models. Inference concerning the coefficient is then said to be be fragile or unreliable, since the coefficient estimate obtained appears to be sensitive to the precise specification of the model used. Some authors have criticized this approach on the grounds of its ad-hoc nature and because it merely presents in a different format the same information as does conventional regression analysis (e.g. Angrist and Pischke (2010), McAleer et al. (1985)).

Model averaging techniques represent an alternative approach to address the issue of fragility in regression analysis. Imagine a situation in which there are many different candidate models for estimating the effect of X on Y. One could estimate a single model (whose selection will probably be based on doubtful assumptions), and then make inference based on that selected model (probably after having a look at inference results from other candidate models previously estimated) ignoring the uncertainty surrounding the selection of that single model. Or, alternatively, one could estimate all the candidate models (or at least a large set of them), and then compute a weighted average of all the estimates for the coefficient on X. Moreover, after computing the associated standard errors, one could make inference based on the whole universe of candidate models. By doing so, we would be considering not only the uncertainty associated to the parameter estimate conditional on a given model, but also the uncertainty of the parameter estimate across different models. This approach would lead us to more reliable, or at least more honest, conclusions regarding the significance of the estimated effect of X on Y. This chapter presents an overview of the literature on model averaging techniques applied to empirical research in economics.

Empirical growth is probably the field in economics in which concerns about fragility of regression estimates have caused the most active literature applying sensitivity analysis techniques. In this context, Levine and Renelt (1992) is possibly the most influential paper considering extreme bound analysis, and Sala-i-Martin et al. (2004) is certainly the paper that popularized the use of model averaging as a tool for applied researchers in the empirical growth regressions literature.

As in any other subfield of econometrics, in this literature there are Bayesian and Frequentist approaches to model averaging. The main differences between Frequentist Model Averaging (FMA) and Bayesian Model Averaging (BMA) arise from the way in which inference is made, and from the selection of model weights. Compared with the FMA approach, there has been a huge literature on the use of BMA in statistics and more recently in economics. Thus, the BMA toolkit is larger than that of FMA. However, the FMA approach is starting to receive a lot of attention over last decade. In this chapter I review the state of the art in both approaches providing a discussion of their advantages and drawbacks. Moreover, I argue that the two perspectives are complementary and they can be reinforced to each other so that dogmatic recommendations are not in the agenda of this chapter.

On the other hand, given the raising interest on causal effects in economics over the last decades, the combination of model averaging and Instrumental Variables (IV) models is an interesting line of open research. Panel data represent an alternative to IV for estimating causal effects in situations with endogenous regressors, so extending the model averaging apparatus to panel data models is also a relevant research topic. The first steps on this direction have been taken during the last lustrum (e.g. Durlauf et al. (2008), Moral-Benito (2009a)). In this chapter I summarize the recent developments on model averaging with endogenous regressors.

The remainder of this chapter is organized as follows. Section 1.2 intuitively presents the basic concepts of model averaging techniques. The Bayesian approach to model averaging is formally presented in Section 1.3, and the Frequentist alternative in Section 1.4. Section 1.5 describes recent model averaging approaches to settings with endogenous regressors, and in Section 1.6 I present some examples of model averaging applications in economics. Finally, Section 1.7 concludes.

1.2 Model Averaging as an (Agnostic) Alternative

It is common in empirical papers to have one baseline specification in which the central conclusions are based on, and several robustness checks in a companion table or even in an appendix. The main objective of this approach is to convince the reader that the results in the paper are robust to different assumptions. However, it is still unclear for the readers how hard the authors had to work to find their favorite outcomes. This situation arises because theory does in general not offer enough guidance in the selection of the appropriate empirical model, and then the empirical researcher faces a problem of model uncertainty. Therefore, different researchers might propose and use different empirical models that are compatible with each other for analyzing the same question. In doing so, they first select a single model (for instance based on a given information criterion or a personal view of the question of interest), and second they make inference under the selected model as if this model had been given in advance. In reality, inference based on the final model might give an excessively optimistic answer due to the under-estimation of the uncertainty associated with the whole estimation procedure. Model averaging represents an alternative to this process.[1] Moreover, I consider model averaging as an agnostic approach in the sense that a researcher employing model averaging techniques is unwilling to commit to an opinion about the best single model.

Frequentist Model Averaging (FMA) and Bayesian Model Averaging (BMA) are two different approaches to model averaging in the literature. Despite their similarities in the spirit and the objectives, both techniques differ in the approach to inference and to the selection of model weights. Sections 3 and 4 present an overview of the developments on BMA and FMA over the last years, and how both approaches conduct inference and compute model weights.

1.2.1 Historical Perspective

As pointed out by Clemen (1989), Laplace (1818) considered combining regression coefficient estimates almost 200 years ago. In particular, he derived and compared the properties of two estimators, one being least squares and the other a kind of weighted median. Moreover, he also analyzed the joint distribution of the two, and proposed a combining formula that resulted in a better estimator than either.[2]

Aside from Laplace, other early treatments of combining multiple estimates came from the statistical literature. Edgerton and Kolbe (1936) propose to combine different estimates in such a way that the combining weights result from minimizing the sum of squares of the differences of the scores. Horst (1938) derives a formula for combining multiple measures in which the criterion is obtaining maximum separation among the individual population members. Halperin (1961) provided a minimum-squared-error combination of estimates, and

[1]Extreme Bound Analysis is also an option to be considered as commented in the Introduction. However, once we accept there are different competing models compatible with each other, model averaging provides a more systematic and rigorous approach to the problem.

[2]See Stigler (1973) for a brief presentation of Laplace's work.

Geisser (1965) appears to be the earliest Bayesian approach to combining estimates. In the forecasting literature, a flood of papers about combining different forecasts was generated in the 1970s since the seminal papers by Barnard (1963) and Bates and Granger (1969). Timmermann (2006) provides a good overview of this literature.

Despite the basic paradigm for Bayesian Model Averaging (BMA) was introduced by Leamer (1978), the approach was basically ignored until the 1990s and 2000s when there has been an enormous literature on the use of BMA. This is so because more powerful computers and theoretical developments such as Markov chain Monte Carlo Model Composition (MC^3) allow researchers to overcome the troubles related to implementing BMA. Raftery (1995) and Fernández et al. (2001b) are good examples of recent BMA applications in economics. The state of research in the field during the nineties was summarized in Hoeting et al. (1999a).

The forecasting combination articles in the 1970s can be considered the predecessors of the current Frequentist Model Averaging (FMA) literature. In contrast to BMA, the FMA approach has started to receive attention over the last decade; see, for example, Hjort and Claeskens (2003) and Hansen (2007). This is so probably because the Frequentist approach to model uncertainty was traditionally focused on model selection rather than model averaging.

1.2.2 Basic Concepts

Empirical research in economics is in general plagued by model uncertainty problems. This means that it is very unlikely that only one model needs to be considered. Imagine a researcher who is trying to estimate the effect of a particular policy on a particular outcome. It is a common situation to have more than one possible model to analyze such effect. Let us suppose that the researcher has q possible models in mind, indexed by $h = 1, ..., q$. This implies that there are q different estimates of the effect of interest depending on the model considered, say $\left\{\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_q\right\}$.

In such a situation, the most common approach is to select a single model from the q existing candidates. There is a huge literature on model selection, i.e. the task of selecting a statistical model from a set of potential models given data. A good overview of this literature can be found in Claeskens and Hjort (2008). After the model selection step, both the inference and the conclusions of the analysis are typically based on this single model. For instance, let us think that the selected model is $h = 3$, and therefore the presented result is $\hat{\beta}_3$. On the other hand, it is also very common to present some extra estimates as a robustness check. It is easy to imagine a final table with several columns presenting some of the estimates corresponding to other possible models, for example $\hat{\beta}_{10}$, $\hat{\beta}_{23}$, $\hat{\beta}_{34}$, and $\hat{\beta}_{50}$. If all the presented estimates are close enough, the researcher concludes that her result is robust and the paper is then ready. However, as previously mentioned, it will be still unclear for the readers how hard the researcher had to work to select and defend the selected model $h = 3$, and more importantly, to find the similar estimates $\hat{\beta}_{10}$, $\hat{\beta}_{23}$, $\hat{\beta}_{34}$, and $\hat{\beta}_{50}$ among the q candidates, bearing in mind that we could easily have billions of candidate models.

Model averaging represents an agnostic alternative to this approach. The key idea of model averaging is to consider and estimate all the q candidate models, and then report a weighted average as the estimate of the effect of interest. Therefore, model averaging is an agnostic approach in the sense that a researcher relying on this approach holds the view that the true single model is unknown and probably unknowable. Then, the best she can do is to consider all the possible alternatives instead of selecting one probably incorrect option. The model averaging estimate ($\hat{\beta}_{MA}$) can then be written as:

$$\hat{\beta}_{MA} = \sum_{h=1}^{q} \omega_h \hat{\beta}_h \tag{1.1}$$

where ω_h represents the weight associated to model h. In subsequent sections we will analyze the different alternatives to choose and estimate the weights ($\hat{\omega}_h$), to obtain the model-specific estimates $\hat{\beta}_h$, and how can we make inference based on model averaging in both its Bayesian and Frequentist versions.

1.3 Bayesian Model Averaging

1.3.1 Estimation and Inference with BMA

For the sake of illustration, let us consider the case of a normal linear regression model in which model uncertainty comes from the selection of regressors to include in the right hand side:

$$\begin{aligned} y &= X\beta + \epsilon \\ \epsilon &\sim N(0, \sigma^2 I_N) \end{aligned} \tag{1.2}$$

where y and ϵ are $N \times 1$ vectors of the dependent variable and the random shocks respectively. X is a $N \times q$ matrix of regressors that may or may not be included in the model, and β ($q \times 1$) contain the parameters to be estimated. If we set some components of $\beta = (\beta_1, \beta_2, ..., \beta_q)'$ to be zeros, there are a total of 2^q candidate models to be estimated —indexed by M_j for $j = 1, ..., 2^q$— which all seek to explain y —the data—. For instance, setting β_1 to be zero implies that we are not including the first regressor (i.e. the first column of X, being $X = (X_1, X_2, ..., X_q)$) in the model. Each model M_j depends upon parameters β^j. In cases where many models are being entertained, it is important to be explicit about which model is under consideration. Hence, following the Bayesian logic, the posterior for the parameters calculated using M_j is written as:

$$g\left(\beta^j | y, M_j\right) = \frac{f\left(y|\beta^j, M_j\right) g\left(\beta^j|M_j\right)}{f\left(y|M_j\right)} \tag{1.3}$$

and now the notation makes clear that we now have a posterior $g\left(\beta^j|y, M_j\right)$, a likelihood $f\left(y|\beta^j, M_j\right)$, and a prior $g\left(\beta^j|M_j\right)$ for each model.

On the other hand, Bayesian inference suggests that the posterior model probability can be used to assess the degree of support for M_j. Therefore, posterior model probabilities will be used as model weights in BMA. Given the prior model probability $P(M_j)$ we can calculate the posterior model probability using Bayes Rule as:

$$P(M_j|y) = \frac{f(y|M_j)\,P(M_j)}{f(y)}. \tag{1.4}$$

According to equations (1.3) and (1.4), it is now clear that we need to elicit priors for the parameters of each model and for the model probability itself. This means that Bayesian Model Averaging (BMA) involves two different prior beliefs, one on the parameter space $(g(\beta^j|M_j))$ and another one on the model space $(P(M_j))$.

In order to calculate the posterior model probability in (1.4) we also need to compute $f(y|M_j)$ that is often called the marginal (or integrated) likelihood, and is calculated using (1.3) and a few simple manipulations. In particular, if we integrate both sides of (1.3) with respect to β^j, use the fact that $\int g(\beta^j|y, M_j)\,d\beta^j = 1$ (since probability density functions integrate to one), and rearrange, we obtain:

$$f(y|M_j) = \int f\left(y|\beta^j, M_j\right) g\left(\beta^j|M_j\right) d\beta^j. \tag{1.5}$$

The quantity $f(y|M_j)$ given by equation (1.5) is the marginal probability of the data, because it is obtained by integrating the joint density of (y, β^j) given y over β^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 β^j are eliminated by maximization rather than by integration.

Following Leamer (1978) we can consider β a function of β^j for each $j = 1, ..., 2^q$ (i.e. $\beta(\beta^j)$) and then calculate the posterior density of the parameters for all the models under consideration by the law of total probability:

$$g(\beta|y) = \sum_{j=1}^{2^q} P(M_j|y)\,g(\beta|y, M_j) \tag{1.6}$$

Therefore, the full posterior distribution of β is a weighted average of its posterior distributions under each of the models, where the weights are given by $P(M_j|y)$.

Given the Bayesian framework based on parameter distributions, when applying BMA according to equation (1.6) both estimation and inference come naturally together from the posterior distribution that provides inference about β that takes full account of model uncertainty.

Despite the Bayesian spirit of the approach, one might also be interested in point estimates and their associated variances. If this is so, one common procedure is to take expectations across (1.6):

$$E(\beta|y) = \sum_{j=1}^{2^q} P(M_j|y)\,E(\beta|y, M_j) \tag{1.7}$$

with associated posterior variance:

$$V(\beta|y) = \sum\nolimits_{j=1}^{2^q} P(M_j|y)\, V(\beta|y, M_j) + \\ + \sum\nolimits_{j=1}^{2^q} P(M_j|y)\, (E(\beta|y, M_j) - E(\beta|y))^2 \tag{1.8}$$

The posterior variance in (1.8) incorporates not only the weighted average of the estimated variances of the individual models but also the weighted variance in estimates of the β'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.

As a by-product of the BMA approach, we can also compute the posterior probability that a particular variable h is included in the regression. In other words, variables with high posterior probabilities of being included are considered as robustly related to the dependent variable of interest. This object 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:

$$\text{posterior inclusion probability} = P(\beta_h \neq 0|y) = \sum\nolimits_{\beta_h \neq 0} P(M_j|y) \tag{1.9}$$

Implementing Bayesian Model Averaging can be difficult because of two reasons: (i) two types of priors (on parameters and on models) need to be elicited and this can be a complicated task. (ii) the number of models under consideration —2^q— is often huge so that the computational burden of BMA can be prohibitive. In the next sections I present some of the remedies proposed in the literature to these problems.

1.3.2 Priors on the Parameter Space

Prior density choice for Bayesian Model Averaging remains an open area of research. In the context of BMA, improper priors for model-specific parameters cannot be used because they are determined only up to an multiplicative arbitrary constant. Despite these constants cancel in the posterior distribution of the model-specific parameters when doing inference for a given model, they remain in marginal likelihoods leading to indeterminate model probabilities and Bayes factors. To avoid this situation, proper priors for β under each model are usually required. Some of the most popular alternatives considered in the literature are summarized below.

Zellner's g Priors

Given the normal regression framework, the bulk of the BMA literature favors the natural-conjugate approach, which puts a conditionally normal prior on coefficients β^j. Virtually all BMA studies use a conditional prior for the j-th model's parameters $(\beta^j|\sigma^2)$ with zero mean and the variance proposed by Zellner (1986), that is, a prior covariance given by

$g(X_j'X_j)^{-1}$. This prior variance is proportional to the posterior covariance arising from the sample $((X_j'X_j)^{-1})$ with the scalar g determining how much importance is attributed to the prior beliefs of the researcher. The conditional prior on β^j is then:

$$\beta^j|\sigma^2, M_j, g \sim N(0, \sigma^2 g(X_j'X_j)^{-1}) \tag{1.10}$$

Moreover, the variance parameter σ is common to all the models under consideration, so an improper prior is not problematic, and the most common approach is the uninformative prior proposed by Fernández et al. (2001a): $p(\sigma) \propto \sigma^{-1}$.[3]

The popularity of this prior structure is due to two factors: (i) it has closed-form solutions for the posterior distributions that drastically reduce the computational burden, and (ii) it only requires the elicitation of one hyperparameter, the scalar g.

Though there are many different options for choosing g (see for example Fernández et al. (2001a)), the three most popular alternatives are:

- Unit Information Prior (g-UIP): proposed by Kass and Wasserman (1995), it corresponds to taking $g = N$, and it leads to Bayes factors that behave like the Bayesian Information Criterion (BIC). Therefore it is possible to combine Frequentist OLS or MLE for estimation with the Schwarz approximation to the marginal likelihood for averaging with a Bayesian justification (see for example Raftery (1995) or Sala-i-Martin et al. (2004)).

- Risk Inflation Criterion (g-RIC): recommended by Foster and George (1994), it implies setting $g = q^2$.

- Benchmark Prior: After a thorough study, Fernández et al. (2001a) determined this combination of the g-UIP and g-RIC priors to perform best with respect to predictive performance. It matches with $g = max(N, q^2)$.

Laplace Priors

Let us construct a partition of the X matrix such that we can rewrite (1.2) as follows:

$$\begin{aligned} y &= X_1\gamma + X_2\delta + \epsilon \\ \epsilon &\sim N(0, \sigma^2 I_N) \end{aligned} \tag{1.11}$$

where γ and δ are the new $q_1 \times 1$ and $q_2 \times 1$ parameter vectors with $q_1 + q_2 = q$.

Given this unrestricted model, we can determine which are the focus regressors (X_1) and which are the auxiliary (doubtful) regressors (X_2).[4] We can reparametrize the model in (1.11) replacing $X_2\delta = X_2^*\delta^*$, with $X_2^* = X_2P\Pi^{-1/2}$ and $\delta^* = \Pi^{1/2}P'\delta$, where P is an orthogonal matrix and Π is a diagonal matrix such that $P'X_2'R_{X_1}X_2P$ and $R_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$.

[3] We can also include a constant term (α) in all the models with prior $p(\alpha) \propto 1$.

[4] Note that the focus regressors may only include a constant term so that we may have the same situation as in the previous section in which all the regressors were focus regressors.

In this setting, Magnus et al. (2010) propose to consider an alternative prior structure that leads to the so-called Weighted-Average Least Squares (WALS) estimator. In particular, WALS use a Laplace distribution with zero mean for the independently and identically distributed elements of the transformed parameter vector $\eta = \delta^*/\sigma$, whose ith element, η_i ($i = 1, ..., q_2$) is the population t-ratio on δ_i, the ith element of δ. As pointed out by Magnus et al. (2010), "this choice of prior moments is based on our idea of ignorance as a situation where we do not know whether the theoretical t-ratio is larger or smaller than one in absolute value".

The WALS estimator employs non-informative model-specific priors and drastically reduces the computational burden of standard BMA being proportional to q_2 (or q) instead of 2^{q_2} (or 2^q). In contrast, WALS does not provide either Bayesian posterior distributions or posterior inclusion probabilities as a measure of robustness.

1.3.3 Priors on the Model Space

In order to implement any of the BMA strategies described above, prior model probabilities ($P(M_j)$) must be assigned. This step might be considered as analogous to the choice of model weights in the Frequentist approach to model averaging (more on this below).

Binomial Priors

For the model size (Ξ), the most common prior structure in BMA research is the Binomial distribution. According to this priors, each variable is independently included (or not) in a model so that model size (Ξ) follows a Binomial distribution with probability of success ξ:

$$\Xi \sim Bin\,(q, \xi) \tag{1.12}$$

where q is the number of regressors considered and ξ is the prior inclusion probability for each variable.

Given the above, the prior probability of a model (M_j) with q_j regressors is given by:

$$P(M_j) = \xi^{q_j}(1-\xi)^{q-q_j} \tag{1.13}$$

One commonly-used particular case of this prior structure is to assume that every model has the same *a priori* probability (i.e. the uniform prior on the model space). This uniform prior corresponds to the assumption that $\xi = 1/2$ so that (1.13) reduces to:

$$P(M_j) = 2^{-q} \tag{1.14}$$

Moreover, given that $E\,(\Xi) = q\xi$, we can fix different priors in terms of both the prior inclusion probability (ξ) or the prior expected model size ($E(\Xi)$). For instance, the uniform prior just described implies $E(\Xi) = q/2$. The choice of one of the hyperparameters ξ or $E(\Xi)$ automatically produces a value for the other, and it leads to larger or smaller penalizations to big models.

Binomial-Beta Priors

Ley and Steel (2009) propose an alternative prior specification in which ξ is treated as random rather than fixed. The proposed hierarchical prior implies a substantial increase in prior uncertainty about model size (Ξ), and makes the choice of prior model probabilities much less critical.

In particular, their proposal is the following:

$$\Xi \sim Bin\left(q,\xi\right) \tag{1.15}$$

$$\xi \sim Be\left(a,b\right) \tag{1.16}$$

where $a,b>0$ are hyper-parameters to be fixed by the researcher. The difference with respect to the Binomial priors is to make ξ random rather than fixed. Model size Ξ will now satisfy:

$$E\left(\Xi\right)=\frac{a}{a+b}q \tag{1.17}$$

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 $b=(q-E(\Xi))/E(\Xi)$ through equation (1.17), so we only need to specify $E(\Xi)$, the prior expected model size, as in the Binomial priors. However, sensitivity of the posteriors with Binomial-Beta priors is smaller than with the Binomial priors.

Dilution Priors

Both the Binomial and the Binomial-Beta priors have in common the implicit assumption that the probability of one regressor appears in the model is independent of the inclusion of others, whereas regressors are typically correlated. In fact, with this priors on model space, a researcher could arbitrarily increase (or reduce) the prior model probabilities across theories simply by including redundant proxy variables for some of these theories. This is the denominated dilution problem raised by George (1999).

To address this issue, Durlauf et al. (2008) introduce a version of George (1999) dilution priors that assigns probability to neighborhoods of models. Moreover, this kind of dilution prior assigns uniform probability to neighborhoods rather than models, and solves the dilution problem. Consider a given theory (or neighborhood of models) (T) for which we have q_T proxies among the whole set of q regressors. For each possible combination of variables corresponding to theory T (C_T) we can assign the following prior probability:

$$P(C_T)=|R_{C_T}|\prod_{h=1}^{q_T}\xi^{\pi_h}(1-\xi)^{1-\pi_h} \tag{1.18}$$

where π_h is an indicator of whether or not variable h is included in the combination C_T and

R_{C_T} is the correlation matrix for the set of variables included in C_T. Since the determinant of this correlation matrix ($|R_{C_T}|$) goes to 1 when the set of variables are orthogonal and to 0 when the variables are collinear, these priors are designed to penalize models with many redundant variables. In practice, we assign the same probability to all the models included in the neighborhood C_T and uniform probability to all the different neighborhoods.

Despite its advantages regarding the dilution property, this prior structure requires agreement on which regressors are proxies for the same theories (i.e. it requires to define the model neighborhoods) which is usually not within reach.

1.3.4 Further Topics in BMA

Computational Aspects

In theory, with the results described above we should be able to carry out BMA. However, in practice, the number of models under consideration (2^q) is often so big that makes it impossible to estimate every possible model. Accordingly, there have been many algorithms developed which carry out BMA without evaluating every possible model.

One possible approach is the so-called Occam's Window proposed by Madigan and Raftery (1994). The basic idea of this technique is to exclude from the summation models that predict the data far less well than the best model, and models that receive less support than any of their simpler submodels. Therefore, using an appropriate search strategy (for instance the leaps and bounds algorithm by Furnival and Wilson (1974)) the number of models to be estimated is drastically reduced. Madigan and Raftery (1994) provide a detailed description of the method.

Another commonly-used alternative, initially developed in Madigan and York (1995) is Markov Chain Monte Carlo Model Composition (MC^3). Markov Chain Monte Carlo (MCMC) methods are common in Bayesian econometrics. MCMC algorithms in general take draws from the parameter space in order to simulate the posterior distribution of interest. However, they do not draw from every region of the parameter space, but focus on regions of high posterior probability. BMA considers the models as discrete random variables so that posterior simulators which draw from the model space instead of the parameter space can be derived. As MCMC in the parameter space, MC^3 takes draws from the model space focusing on models with high posterior model probability. Implementing and programming MC^3 is very intuitive and it is not complicated. In the Appendix you can find a detailed description of how does MC^3 work in practice.

A Frequentist Approach to BMA?

If we assume diffuse priors on the parameter space for any given sample size, or, if we have a large sample for any given prior on the parameter space we can write equation (1.7) as follows:[5]

$$E(\beta|y) = \sum_{j=1}^{2^q} P\left(M_j|y\right) E(\beta|y, M_j) = \sum_{j=1}^{2^q} P\left(M_j|y\right) \widehat{\beta}_{ML}^{j} \tag{1.19}$$

where $\widehat{\beta}_{ML}^{j}$ is the ML estimate for model j.

If one is interested in model averaged point estimates, we can use the Schwarz asymptotic approximation to the Bayes factor and uniform model priors so that:

$$P\left(M_j|y\right) = \frac{f(y|\widehat{\beta}_j, M_j) N^{\frac{-q_j}{2}}}{\sum_{i=1}^{2^q} f(y|\widehat{\beta}_i, M_i) N^{\frac{-q_i}{2}}} \tag{1.20}$$

where $f(y|\widehat{\beta}_j, M_j)$ is the maximized likelihood function for model j.

Comparing this expression with Frequentist model weights based on information criteria (see Section 1.4.3), and given the use of maximum likelihood estimates, I argue that this commonly-used approach to BMA (e.g. Raftery (1995), Sala-i-Martin et al. (2004), Moral-Benito (2009a)) can be considered as a Frequentist BMA method.

This approach was first proposed by Raftery (1995) in a general setting. Sala-i-Martin et al. (2004) popularized its use in economics averaging model-specific OLS estimates in the so-called Bayesian Averaging of Classical Estimates (BACE). Finally, Moral-Benito (2009a) generalized the use of this approach to panel data models in the denominated Bayesian Averaging of Maximum Likelihood Estimates (BAMLE).

Moreover, as noted by Moral-Benito (2009b), posterior distributions of the parameters can also be obtained with this approach. Analogously to the posterior mean, these posterior distributions are weighted averages of marginal posterior distributions conditional on each individual model. More concretely, these posteriors are mixture normal distributions because model-specific posteriors are normal. This is so because we can make use of the Bernstein-von Mises theorem[6] (also known as the Bayesian CLT) which basically states that a Bayesian posterior distribution is well approximated by a normal distribution with mean at the MLE and dispersion matrix equal to the inverse of the Fisher information.

1.4 Frequentist Model Averaging

1.4.1 Definition of FMA Estimators

Let us take the linear model in matrix form to illustrate the definition of the FMA estimator:

$$y = \beta X_1 + X_2 \gamma + U \tag{1.21}$$

[5]The equivalence of classical inference and Bayesian inference under diffuse priors is well-known in the classical normal regression model. For the LIML case, Kleibergen and Zivot (2003) show this equivalence for a particular choice of non-informative priors. Note also that the large sample equivalence is only an approximation.

[6]Berger (1985) provides an in-depth analysis and an excellent illustration.

where y, X_1, and U are $N \times 1$ vectors of the dependent variable, the treatment variable of interest and the random shocks respectively. X_2 is a $N \times q$ matrix of doubtful control variables that may or may not be included in the model, and β and γ $(q \times 1)$ contain the parameters to be estimated. Despite we make this distinction between X_1 and X_2 for illustration purposes, FMA can easily handle situations in which we cannot make such a distinction. Finally, N is the number of observations in the sample.

If we set some components of $\gamma = (\gamma_1, \gamma_2, ..., \gamma_q)'$ to be zeros, there are a total of 2^q candidate models to be estimated. Given the coefficient of interest is β, let $\hat{\beta}_M$ be the estimator of β under the candidate model M with $M \in \{M_1, M_2, ..., M_{2^q}\}$. The most common approach in applied research is to take the selected model as given and base the inference on this single estimate $\hat{\beta}_M$ while the actual estimator is:

$$\hat{\beta} = \begin{cases} \hat{\beta}_{M_1} & \text{if the first model is selected} \\ \hat{\beta}_{M_2} & \text{if the second model is selected} \\ \vdots & \vdots \\ \hat{\beta}_{M_{2^q}} & \text{if the } 2^q\text{-th model is selected} \end{cases}$$

We can also rewrite the above estimator as

$$\hat{\beta} = \sum_{j=1}^{2^q} \tilde{\omega}_{M_j} \hat{\beta}_{M_j}$$

where:

$$\tilde{\omega}_{M_j} = \begin{cases} 1 & \text{if the candidate model } M_j \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

This estimator is the usual pre-test estimator that suffers from the previously commented drawbacks if model uncertainty (in the selection of the control variables for example) is present. Therefore, we consider the smoothed weights ω_{M_j} and accordingly, the FMA estimator is given by:

$$\hat{\beta}_{FMA} = \sum_{j=1}^{2^q} \omega_{M_j} \hat{\beta}_{M_j} \tag{1.22}$$

where $0 \leq \omega_{M_j} \leq 1$, and $\sum_{j=1}^{2^q} \omega_{M_j} = 1$. Such estimator is labeled as the FMA estimator of β which integrates the model selection and estimation procedure.

1.4.2 FMA Inference

Hjort and Claeskens (2003) studied the asymptotic properties of the FMA estimator with the form in equation (1.22). The main result is the obtaining of its asymptotic distribution:[7]

$$\sqrt{N}\left(\hat{\beta}_{FMA} - \beta_{true}\right) \xrightarrow{d} \Lambda \tag{1.23}$$

[7] More details about the derivation of this distribution can be found in the Appendix.

where $\Lambda = \sum_{j=1}^{2^q} \omega_{M_j} \Lambda_j$.

However, inference based on this limiting distribution Λ will still ignore the uncertainty involved in the model selection step. Therefore, confidence intervals constructed from $\hat{\beta}_{FMA}$ and the variance of Λ in the usual way, will produce too optimistic inference and will lead to misleading conclusions because the real coverage probability is lower than the intended level.

In response to this problem, Buckland et al. (1997) proposed an alternative approach to deal with this issue when constructing confidence intervals of FMA estimators. Their method takes the extra model uncertainty into account by including an extra term in the variance of the FMA estimator. In particular, the proposed formula for the estimated standard error of $\hat{\beta}_{FMA}$ is:

$$\widehat{SE}(\hat{\beta}_{FMA}) = \sum_{j=1}^{2^q} \omega_{M_j} \sqrt{\hat{\tau}_j^2/N + \hat{b}_j^2} \tag{1.24}$$

where $\hat{\tau}_j^2$ estimates the variance of Λ_j, and $\hat{b}_j = \hat{\beta}_{M_j} - \hat{\beta}_{FMA}$ captures the extra uncertainty associated with the variation of estimates across different models. This formula implies an estimated variance for the FMA estimator that closely resembles its Bayesian counterpart in equation (1.8) (i.e. the posterior variance for the BMA estimator). Note that we still have to replace the fixed weights in equations (1.22) and (1.24) by their estimates in order to apply FMA.

1.4.3 Model Weights in FMA

FMA estimators crucially depend on the weights selected for estimation. In the previous subsections the weights were taken as fixed, but it is important to remark here that different weights will result in different asymptotic properties of the corresponding FMA estimators.

Weight Choice Based on Information Criteria

Probably the most common approach to weight choice in Frequentist Model Averaging is the one based on different information criteria of the form:

$$I_j = -2\log(L_j) + \varphi_j$$

where L_j is the maximized likelihood function for the j-th model, and φ_j is a penalty term function of the number of parameters and/or the number of observations of model j (i.e. q_j).

Buckland et al. (1997) propose to use the following model weights:

$$\omega_{M_j} = \frac{exp(-I_j/2)}{\sum_{h=1}^{2^q} exp(-I_h/2)} \tag{1.25}$$

The penalty term $\varphi_j = 2q_j$ corresponds to the Akaike Information Criterion (AIC), being q_j the number of parameters in model j. Therefore Akaike weights are one common alternative. Another possible choice is $\varphi_j = q_j \ln(N)$ that corresponds to the Bayesian Information

Criterion (BIC). Given the use of BIC is also justified from a Bayesian viewpoint, this illustrates one clear similarity between BMA and FMA, that provide in fact the same point estimates under some particular circumstances.

Information criteria such as the AIC and the BIC select one single best model regardless of the parameter of interest. However, there are situations in which one model is best for estimating one parameter, whereas another model is best for another parameter. Aware of this situation, Claeskens and Hjort (2003) propose to use the Focused Information Criterion (FIC) to select the best model, but depending on the parameter of interest. Of course, the FIC can naturally be employed as an alternative to construct FMA model weights.

Weight Choice Based on Mallow's Criterion

Hansen (2007) proposes to select the model weights in least squares model averaging by minimizing the Mallow's criterion. Despite this criterion is similar to the Akaike information criterion in the model selection spirit, the approach to calculate the weights in Hansen (2007) in the model averaging setting is different.

Hansen (2007) considers the following homoskedastic linear regression:

$$\begin{aligned} y_i &= \sum_{j=1}^{\infty} \theta_j x_{ij} + e_i \\ E(e_i|x_i) &= 0 \\ E(e_i^2|x_i) &= \sigma^2 \end{aligned} \tag{1.26}$$

where $x_i = (x_{i1}, x_{i2}, ...)$.

Now consider the sequence of candidate models $j = 1, 2, ...$ seeking to approximate (1.26). The j-th model uses the first ϕ_j elements of x_i with $0 < \phi_1 < \phi_2 < ....$ Given the above, the j-th candidate model is:

$$y_i = \sum_{j=1}^{\phi_j} \theta_j x_{ij} + e_i \tag{1.27}$$

with corresponding approximating error $\sum_{j=\phi_j+1}^{\infty} \theta_j x_{ij}$. Let us rewrite (1.27) in matrix form:

$$Y = X_j \Theta_j + e \tag{1.28}$$

where Y and e are $N \times 1$ vectors, X_j is a $N \times \phi_j$ matrix, and Θ_j is a $\phi_j \times 1$ vector of parameters. Let $J = J(N) \leq N$ be the candidate model with the largest number of regressors, and $\lambda = (\lambda_1, ..., \lambda_J)'$ a weight vector in the unit simplex in $\mathbb{R}^J$:

$$\mathbf{H}_N = \left\{ \lambda \in [0, 1]^J : \sum_{j=1}^{J} \lambda_j = 1 \right\}$$

The least squares model averaging estimator of Θ_J can be defined as:

$$\hat{\Theta}_J(\lambda) = \sum_{j=1}^{J} \lambda_j \begin{pmatrix} \hat{\Theta}_j \\ 0 \end{pmatrix}$$

where $\hat{\Theta}_j$ represents the least squares estimate of model j.

We are now ready to introduce the Mallows' criterion to be minimized in order to obtain the model weights:

$$\hat{\lambda} = \underset{\lambda \in \mathbf{H}_N}{\text{argmin}} \; C_N(\lambda)$$

where:

$$C_N(\lambda) = (Y - X_J\hat{\Theta}_J(\lambda))'(Y - X_J\hat{\Theta}_J(\lambda)) + 2\sigma^2\lambda'\Phi$$

with $\Phi = (\phi_1, ..., \phi_J)'$.

Furthermore, Hansen (2007) provides an optimality result of his Mallows Model Averaging (MMA) estimator. In particular, it states that the MMA estimator achieves the lowest possible squared error when we constrain the weight vector to the discrete set $\mathbf{H}_N$ (i.e. it is asymptotically optimal). However, it is important to mention that the optimality of MMA fails under heteroskedasticity.

In a situation of instrument uncertainty (i.e. many candidate instruments for a given set of endogenous variables), Kuersteiner and Okui (2010) propose to apply the MMA approach to the first stage of the 2SLS, LIML and Fuller estimators, and then use the average predicted value of the endogenous variables in the second stage.

Weight Choice Based on Cross-Validation Criterion

In a recent paper, Hansen and Racine (2009) propose how to optimally average across non-nested and heteroskedastic models. In particular, they suggest to select the weights of the least squares model averaging estimator by minimizing a deleted-1 cross-validation criterion, so that the approach is labeled as Jackknife Model Averaging (JMA). In comparison with MMA, JMA (and its optimality result) is appropriate for more general linear models (i.e. random errors may have heteroskedastic variances, and the candidate models are allowed to be non-nested). Aside from this two points, the setup is the same as for the MMA estimator in (1.26). Let us further define:

$$\mu_i = \sum_{j=1}^{\infty} \theta_j x_{ij}$$

so that the jackknife version of the model averaging estimator of μ is:

$$\hat{\mu}(\lambda) = \sum_{j=1}^{J} \lambda_j \hat{P}_j Y = \hat{P}(\lambda)Y$$

where $\hat{P}_j = \hat{D}_j(P_j - I_N) + I_N$, $P_j = X_j(X_j'X_j)^{-1}X_j'$ is the projection matrix under the j-th candidate model, $\hat{D}_j$ is the $N \times N$ diagonal matrix with the i-th diagonal element being

$(1-h_{ii}^{j})^{-1}$, $h_{ii}^{j} = X_{j,i}(X_j'X_j)^{-1}X_{j,i}'$, and $X_{j,i}$ is the i-th row of X_j. The deleted-1 cross-validation criterion is defined as:

$$CV(\lambda) = (Y - \hat{\mu}(\lambda))'(Y - \hat{\mu}(\lambda))$$

Finally, the JMA estimator is $\hat{\mu}(\hat{\lambda}^*)$ with weights given by:

$$\hat{\lambda}^* = \underset{\lambda \in \mathbf{H}_N}{\operatorname{argmin}}\ CV(\lambda)$$

Moreover, there is also a theorem that builds the asymptotic optimality of the JMA estimator in the sense of achieving the lowest possible expected squared error. Hansen and Racine (2009) also conduct Monte Carlo simulations showing that JMA can achieve significant efficiency gains over existing model selection and averaging methods in the presence of heteroskedasticity.

1.5 Model Averaging and Endogeneity

The methods described above are all based on the strict exogeneity assumption of the regressors. This assumption implies that there is no correlation between the Xs and the unobservables (ϵ) affecting the output Y (i.e. $cov(X, \epsilon) = 0$). In the treatment effects terminology, it corresponds to the assumption that the treatment is conditionally randomly assigned to the population so that the ordinary least squares (OLS) estimates of the parameters can be interpreted as causal effects. However, in many applications such as empirical growth regressions this assumption is clearly violated. Therefore we might have some Xs, say X_1, that are endogenous, and some others, say X_2, that are exogenous (i.e. X_1 variables are correlated with the unobservables given the X_2 variables and thus $cov(X_1, \epsilon|X_2) \neq 0$). Under these circumstances, obtaining estimates of causal effects requires the availability of an exogenous source of variation on the endogenous variables, that is, a set of valid instruments (Z) which satisfies the conditional IV identifying assumption $cov(Z, \epsilon|X_2) = 0$. Given the interest on causal effects over the last decades, how to tackle the issue of endogeneity in the model averaging framework is an important line of open research.

Formally, when we face a situation in which we have endogenous (X_1) and exogenous (X_2) regressors together with a set of valid instruments (Z) in a linear context, the model to be estimated is:

$$\begin{aligned} y &= X_1\beta_1 + X_2\beta_2 + \epsilon \\ X_1 &= Z\pi_1 + X_2\pi_2 + V \end{aligned} \tag{1.29}$$

where y and X_1 are the $N \times 1$ vector and the $N \times q_1$ matrix of endogenous variables, X_2 is the $N \times q_2$ matrix of exogenous regressors or control variables,[8] and Z corresponds to the $N \times q_Z$

[8] We can also refer to the exogenous regressors X_2 as control or conditioning variables in the sense that, in some cases, they must be included in the model in order to guarantee the validity of the instruments even

matrix of instrumental variables. Moreover, β_1, β_2, π_1 and π_2 represent the $q_1 \times 1$, $q_2 \times 1$, $q_Z \times q_1$ and $q_2 \times q_1$ vectors and matrices of parameters respectively. Finally, the unobservables in the first equation (i.e. the structural form equation) are captured by the $N \times 1$ vector ϵ, and V is the $N \times q_1$ matrix of errors corresponding to the q_1 remaining equations (usually labeled as reduced form equations).

In this framework, we can define the $Q \times 1$ vector $U_i = (\epsilon_i, V_i')'$ and further assume:

$$U_i \sim N(0, \Sigma) \tag{1.30}$$

where Σ is a $Q \times Q$ symmetric and positive definite covariance matrix, and $Q = 1 + q_1$. Given this assumption we can construct the (pseudo) likelihood function for such a model and estimate the parameters via (pseudo) maximum likelihood (i.e. Limited Information Maximum Likelihood (LIML)), or we can estimate the parameters via two-stage least squares (2SLS). In both cases we need to have as many instruments as endogenous regressors ($q_Z \geq q_1$) together with the rank condition $rank(E(Z'X_1)) = q_1$ in order to guarantee identification. In the just-identified case ($q_Z = q_1$), LIML and 2SLS coincide.

Given the IV setting described above, two main sources of model uncertainty arise. In particular we might have uncertainty surrounding the selection of endogenous variables X_1 of interest, and uncertainty in the choice of exogenous (or control) variables X_2. As previously stated, how to address the problem of model uncertainty in these settings is an open issue in the model averaging literature. Given the LIML likelihood function and following techniques advanced by Raftery (1995), one natural possibility is the combination of LIML estimates with BIC model weights. An important remark here is the importance of considering comparable likelihoods across models. Even in the case of a model not including some elements of X_1 in the structural equation, for the sake of comparability we need to consider the full set of reduced form equations for all the variables in X_1. This means that we must construct for this model the likelihood $f(y, X_1|X_2, Z)$ with the full set of candidate endogenous variables in order to guarantee comparability with all the other models under consideration, i.e., the joint likelihood of y and X_1 must be constructed for all the models. The differences across models emerge in the form of zero restrictions on the parameter vectors β_1, β_2, and π_2 for those variables (either X_1 for β_1, or X_2 for β_2 and π_2) not included in a particular model. However, the key point is that the set of q_1 reduced form equations for X_1 must be considered in all the candidate models (i.e. the matrix π_1 is the same in all the models) despite not all the q_1 endogenous variables in X_1 are included in all the models' structural form equations given the existence of model uncertainty in the choice of such variables.

In order to present the LIML likelihood, note that the model in (1.29) can be written as follows:

$$\left(\begin{array}{c:c} 1 & -\beta_1' \\ \hdashline 0 & I_{q_1} \end{array}\right) \left(\begin{array}{c} y' \\ \hdashline X_1' \end{array}\right) = \left(\begin{array}{c:c} \beta_2' & 0 \\ \hdashline \pi_2' & \pi_1' \end{array}\right) \left(\begin{array}{c} X_2' \\ \hdashline Z' \end{array}\right) + \left(\begin{array}{c} \epsilon' \\ \hdashline V' \end{array}\right) \tag{1.31}$$

if their effect is not of central interest.

or more compactly:

$$BY' = CW' + U' \tag{1.32}$$

where:

$$B = \left(\begin{array}{c|c} 1 & -\beta_1' \\ \hline 0 & I_{q_1} \end{array}\right)_{Q\times Q}$$

$$Y' = \left(\begin{array}{c} y' \\ \hline X_1' \end{array}\right)_{Q\times N}$$

$$C = \left(\begin{array}{c|c} \beta_2' & 0 \\ \hline \pi_2' & \pi_1' \end{array}\right)_{Q\times(q_2+q_Z)}$$

$$W' = \left(\begin{array}{c} X_2' \\ \hline Z' \end{array}\right)_{(q_2+q_Z)\times N}$$

$$U' = \left(\begin{array}{c} \epsilon' \\ \hline V' \end{array}\right)_{Q\times N}$$

The gaussian log-likelihood function of the full model (i.e. the model that includes all the candidate variables) is:

$$\ln f(Y|W, M_f) = -\frac{NQ}{2}\log 2\pi + N\log|\det B| - \frac{N}{2}\log\det\Sigma - \frac{1}{2}tr(\Sigma^{-1}U'U) \tag{1.33}$$

where note that Y includes both y and X_1, W includes X_2 and Z, and M_f refers to the full model including all the candidate variables available. Despite the fact that number of parameters to be estimated might be huge and the problem might become computationally unfeasible from a model averaging perspective, we can concentrate out the reduced form parameters and drastically reduce the computational burden (see Moral-Benito (2009b)).

Having the likelihood function for the full model, it is easy to obtain the likelihood functions for the remaining models in order to compute the marginal likelihoods and model weights (or posterior model probabilities). Given the focus on model averaging, we need all the model-specific likelihoods in which some parameters are restricted to be zero depending on the variables (either endogenous or exogenous) included in the model. If a given endogenous variable is excluded from the full model, we simply restrict to zero the corresponding element of the β_1 vector of coefficients, but the rest of the likelihood remains unchanged in order to guarantee comparability across models. If the excluded variable is an exogenous one, we restrict to zero the corresponding elements of the vectors β_2 and π_2.[9]

As advanced by Raftery (1995), we can extend the model averaging approach to the setting of endogenous variables by computing BIC weights from the LIML likelihoods just

[9] Note that this particular likelihoods with restrictions would not be necessary if we are not interested in model comparison and our only interest is the estimation of a single model.

described. Given the likelihood function, Section 1.3.4 of this chapter formally presents the approach and its Bayesian justification. Moreover, this is also the approach considered by Moral-Benito (2009a) and Moral-Benito (2009b) in a panel data setting (see below for more details).

In the cross-sectional setting, Durlauf et al. (2008) represents the first attempt to address the issue of endogenous regressors in a BMA context.[10] More concretely, the paper is concerned with uncertainty surrounding the selection of the endogenous and exogenous variables of interest. Therefore they consider $2^{q_1+q_2}$ candidate models indexed by $j = 1, ..., 2^{q_1+q_2}$. The authors propose to use 2SLS model-specific estimates for each single model, and then take the average:

$$E(\theta|y) = \sum_{j=1}^{2^{q_1+q_2}} P\left(M_j|y\right) E(\theta|y, M_j) \approx \sum_{j=1}^{2^{q_1+q_2}} P\left(M_j|y\right) \widehat{\theta}^j_{2SLS} \tag{1.34}$$

where $\widehat{\theta}^j_{2SLS}$ is the 2SLS estimate for model j, and $\theta = (\beta_1, \beta_2, vec(\pi_1), vec(\pi_2), vech(\Sigma))$ is the $h \times 1$ vector of parameters to be estimated.

The weights $P\left(M_j|y\right) \propto f(y|M_j)P(M_j)$ are inspired in a limited information version of the BIC (i.e. LIBIC) approximation to the integrated likelihood $f(y|M_j)$:

$$f(y|M_j) \approx \exp\big[-\frac{N(q_1+1)}{2}\log(2\pi) - \frac{N}{2}\log(\det(\hat{\Sigma})) - \frac{h}{2}\log N\big] \tag{1.35}$$

where $\hat{\Sigma} = \sum_{i=1}^{N} \hat{U}_i \hat{U}_i'$ and $\hat{U}_i$ is the predicted residual from the 2SLS estimates. With respect to model priors ($P(M_j)$), Durlauf et al. (2008) use the dilution priors previously described.

The main drawback of this approach is that its formal justification remains an open issue as stated by the authors. They only give an heuristic interpretation to the results emerging from this averaging of 2SLS estimates. Moreover, since the model weights are based on "pseudo" likelihoods that are in principle not comparable across models, the comparability of these model weights represents another important caveat of the approach. This is so because in this approach, if an endogenous variable is not included in the model, its associated reduced form equations are not considered either. Then, the different models have "pseudo" likelihoods functions for different variables which are not comparable (e.g. the likelihood $f(y, x_1|z)$ is not comparable to the likelihood $f(y, x_2|z)$). More recently, Durlauf et al. (2009) consider model averaging across just-identified models so that model-specific 2SLS estimates $\widehat{\theta}^j_{2SLS}$ coincide with model-specific LIML estimates $\widehat{\theta}^j_{LIML}$ so that the proper likelihood-based BIC weights have formal justification but are still not comparable across models.

In a recent paper, Eicher et al. (2009a) extend BMA to formally account for model uncertainty not only in the selection of endogenous and exogenous variables, but also in the selection of instruments (Z). This third source of uncertainty emerges if we have a set of instruments that satisfy all the exclusion restrictions given a set of endogenous variables regardless of the particular model considered and thus we do not know which instruments

[10] Despite the section is devoted to the connection between model averaging and endogeneity, all advances on this direction are based on the Bayesian spirit of model averaging.

to include in a given model. In the previous setting, the inclusion of a given endogenous variable in the model implied its own set of valid instruments to be included in the model. In particular, Eicher et al. (2009a) propose a 2-step procedure that first averages across the first-stage models (i.e. linear regressions of the endogenous variables on the exogenous ones and the instruments) and then, given the fitted endogenous regressors from the first stage it again takes averages in the second stage. In both steps the authors propose to use BIC weights. In this approach, some issues such as the selection of instruments (i.e. identification of some particular models which depends on q_1 and q_Z) and the statistical justification for the use of BIC-inspired weights in this context are still unclear.

All in all, given the potential criticisms of the approaches to model averaging and endogeneity in the existing literature (e.g. Durlauf et al. (2008), Eicher et al. (2009a)), the approach from first principles presented at the beginning of this section seems to be the most appropriate one to simultaneously address the issues of model uncertainty and endogeneity. The proposed approach, based on averaging across LIML estimates, guarantees comparability of the model likelihoods, and it has statistical justification (see Raftery (1995)). Its main disadvantage is the greater computational burden due to the large number of reduced form parameters to be estimated in all the models under consideration.[11]

1.5.1 Panel Data and Model Averaging

Another relevant open line of research is that of model averaging and panel data as an alternative approach to address the issue of endogenous regressors. Omitted variables biases arising from individual-specific and time-invariant unobservable factors can be alleviated by resorting to panel data models with fixed effects. Panel data comprises information on individuals ($i = 1, ..., N$) over different time periods ($t = 1, ..., T$). Therefore, the correlation between X and ϵ might arise because of a time-invariant and individual-specific characteristic (η_i), that is a component of ϵ_i ($\epsilon_i = \eta_i + \vartheta_{it}$) so that:

$$y_{it} = x_{it}\beta + \eta_i + \vartheta_{it} \tag{1.36}$$

where:

$$E(\eta_i|x_i) \neq 0 \tag{1.37}$$
$$E(\vartheta_{it}|x_i, \eta_i) = 0 \tag{1.38}$$

where x_i is a $T \times 1$ vector such that $x_i = (x_{i1}, x_{i2}, ..., x_{iT})'$.

Assumption (1.37) indicates that the regressors are correlated with the time invariant component of the error term and thus they are endogenous with respect to the fixed effects. However, assumption (1.38) is usually labeled as a strict exogeneity assumption that represents the independence of the transitory shocks (ϑ) with respect to both the regressors and

[11] Note however that concentration of the likelihood functions with respect to the common parameters reduces this computational burden (e.g. Moral-Benito (2009b)).

the permanent component of the shock (η).

Moral-Benito (2009a) considers such a panel setting and shows how to combine different panel data estimators with BMA techniques using different prior structures. In particular, Moral-Benito (2009a) extends to the panel data setting in (1.36)-(1.38) the Benchmark g-Priors and a diffuse prior (in the spirit of Sala-i-Martin et al. (2004)) on the parameter space, and the Binomial and Binomial-Beta priors on the model space. For the model weights, he considers BIC inspired weights that have both a Bayesian (g-UIP prior) and a Frequentist (Schwarz) interpretation.

In this framework we might also have dynamics that complicate the model-specific estimation step. In particular, the vector x_{it} can also include a lagged dependent variable (y_{it-1}) which is correlated with ϑ_{it-1} by definition and thus assumption (1.38) is violated. Moral-Benito (2009a) also considers a dynamic model as follows:

$$y_{it} = \alpha y_{it-1} + x_{it}\beta + \eta_i + \vartheta_{it} \tag{1.39}$$

where:

$$E(\eta_i|y_i, x_i) \neq 0 \tag{1.40}$$
$$E(\vartheta_{it}|y_i^{t-1}, x_i, \eta_i) = 0 \tag{1.41}$$

where y_i^{t-1} is a $(t-1)\times 1$ vector such that $y_i^{t-1} = (y_{i1}, y_{i2}, ..., y_{it-1})'$. Assumption (1.41) makes it clear that endogeneity generated by the dynamics of the model is taken into account.

In this setting, uncertainty comes from the selection of the xs to be included in the model. Moral-Benito (2009a) proposes to combine BMA with a panel likelihood-based estimator which allows obtaining consistent estimates of the autoregressive parameter α based on assumptions (1.40)-(1.41). Moreover, since model-specific estimates are based on a proper likelihood function, model weights are given by the BIC approximation to the marginal likelihood (see Raftery (1995)). The approach is labeled Bayesian Averaging of Maximum Likelihood Estimates (BAMLE).

On the other hand, panel data can also be useful for addressing the biases arising from reverse causality, which is a source of bias different from the biases just described. Coming back to the static setting in (1.36), the reverse causality problem arises if, instead of the assumptions in (1.37)-(1.38), we face:

$$\begin{aligned} E(\eta_i|x_i) &\neq 0 \\ E(\vartheta_{it}|x_i, \eta_i) &\neq 0 \end{aligned} \tag{1.42}$$

In this setting, using panel data without additional instruments (Z) one can obtain causal effect estimates based on the following identification strategy: realizations of the endogenous regressors far enough in time are independent of the current shocks[12] (i.e. $cov(x_{it-\tau}, \vartheta_{it}) = 0$).

[12]Note that this is the same strategy as the one adopted in the dynamic panel setting above in which one could interpret that the lagged dependent variable is an additional endogenous regressor.

Then, we can use this previous realizations ($x_{it-\tau}$) as "internal" instruments in the spirit of (1.29). Using this strategy, Moral-Benito (2009b) constructs a likelihood function for panel data models with unobserved heterogeneity (i.e. fixed effects) and endogenous regressors and combines this likelihood with BMA methods employing the BAMLE approach.

The same panel setting is also considered in Chen et al. (2009) who combine panel GMM estimators with BMA. In particular, they interpret the exponentiated GMM objective function as the model-specific "pseudo" marginal likelihood, and then use LIBIC weights in the spirit of Durlauf et al. (2008). The formal justification of the considered model-specific "pseudo" likelihoods is an important caveat of this approach as in the cross-sectional case discussed in the previous subsection.[13]

1.6 Model Averaging in Economics

Until the 1990s, the bulk of the literature on model averaging came from two different sources: on the one hand, statistical papers developing the BMA apparatus with little emphasis on economic applications (e.g. Raftery (1995), Fernández et al. (2002), Volinsky et al. (1997)), and, on the other hand, papers from the forecasting combination literature to be discussed below. However, since the beginning of the 21^{st} century, new methods together with more powerful computers are inspiring a flurry of BMA activity in different fields of economics.

Empirical growth is, without any doubt, the most active field in which model averaging techniques are being applied. In the search for a satisfactory empirical model of growth, the main area of effort has been the selection of appropriate variables to include in linear growth regressions. The literature concerned with this task is enormous: a huge number of papers have claimed to have found one or more variables correlated with the growth rate, resulting in a total of more than 140 variables proposed as growth determinants. However, given the limited number of observations, the fragility of these regressions causes a big concern among growth researchers.

In an attempt to investigate the robustness of the results, Levine and Renelt (1992) employ the extreme-bounds analysis proposed by Leamer (1983) and Leamer and Leonard (1983), and they concluded that very few variables (e.g. investment) were robustly correlated with growth. In contrast, Sala-i-Martin (1997) relaxed the robustness requirements, constructed weighted averages of OLS coefficients and found that some were fairly stable across specifications. However, the way in which standard errors and distribution of estimates are computed in Sala-i-Martin (1997) are somehow ad hoc and they lack formal justification. The seminal papers on model averaging and growth are Fernández et al. (2001b) and Sala-i-Martin et al. (2004). Following Raftery (1995), Sala-i-Martin et al. (2004) combine OLS estimates with BIC weights in a pseudo-Bayesian approach denominated Bayesian Averaging of Classical Estimates (BACE). On the other hand, Fernández et al. (2001b) employ the Benchmark g Priors for the parameters in a pure Bayesian spirit. Both approaches use the Binomial

[13] It is also worth mentioning the well-documented finite sample biases in the panel GMM estimates they consider (see for instance Alvarez and Arellano (2003)).

prior on the model space with different prior expected model sizes. Magnus et al. (2010) consider the WALS approach with uniform model priors in the growth context, and Wagner and Hlouskova (2009) apply a FMA estimator based on principal components using four weighting schemes: equal, MMA, AIC, and BIC. Durlauf et al. (2008) is the first paper worried about causal effects in BMA empirical growth research using BIC weights and dilution priors on the model space. Moral-Benito (2009a) extends the use of model averaging techniques to a panel data setting considering and comparing different prior structures on both the parameter space and the model space. In the spirit of Raftery (1995), Moral-Benito (2009a) proposes to combine maximum likelihood estimates with model averaging using BIC weights in the so-called Bayesian Averaging of Maximum Likelihood Estimates (BAMLE) approach. Finally, Moral-Benito (2009b) and Mirestean and Tsangarides (2009) combine model averaging techniques with causal effect estimates in a panel data framework.

Since the seminal paper by Bates and Granger (1969), there has been an enormous literature on forecast combination with the aim of improving forecasting performance. From a Frequentist viewpoint, there is a vast empirical literature on forecast combining, and there are also a number of simulation studies that compare the performance of combining methods in controlled experiments. These studies are surveyed by Diebold and Lopez (1996) and Timmermann (2006). With respect to the Bayesian approach to model averaging, there are many BMA applications to forecasting financial variables such as stock returns (e.g. Avramov (2002), Cremers (2002)) or exchange rates (e.g. Wright (2008a)). In the macro forecasting literature, Garratt et al. (2003) employ BMA to predict inflation and output growth in the UK and Wright (2008b) forecasts US inflation by BMA.

Aside from empirical growth and forecast combination, model averaging techniques are becoming popular in other fields of economics. Pesaran et al. (2009) employ model averaging as a remedy to the risk of inadvertently using false models in portfolio management; Wan and Zhang (2009) consider FMA estimators to determine the degree to which recreation and tourism development affected a range of socioeconomic indicators (e.g. earnings per job, income per capita, etc.) in 311 rural U.S. counties in the 1990's and 2000; in labor economics, Tobias and Li (2004) apply model averaging to estimate Mincer equations; Cohen-Cole et al. (2007) study the controversial issue of the deterrent effect of capital punishment employing a BMA approach; Galbraith and Hodgson (2009) analyze the determinants of the value of works of art using model averaging.

1.7 Conclusions and Future Research

The choice of control variables in regression analysis applied to empirical research in economics is a critical issue that is, in general, underestimated. As illustrated by Leamer (1983) among others, conclusions from empirical studies may well depend on the controls included, so that the results are sensitive to different choices of control explanatory variables. Model averaging approaches estimate the effect of interest under all the possible combinations of controls, and report a weighted average effect. Therefore, model averaging takes into account the uncertainty surrounding the selection of controls (i.e. model uncertainty) in a natural manner.

This chapter has presented an overview of existent model averaging techniques and their applications in economics. Both the Frequentist and the Bayesian approaches to model averaging have been summarized. Bayesian Model Averaging (BMA) involves the elicitation of model and parameter priors; Frequentist Model Averaging (FMA) requires to choose model weights and model-specific estimators. Several alternatives on both sides have been described in this chapter. Moreover, an attempt to connect both approaches is made in Section 1.3.

How to tackle the issue of endogenous regressors in the model averaging framework is an interesting line of open research. The state of the art of the literature on BMA and endogeneity in the conditional IV and panel settings has been summarized in this chapter. Allowing for endogenous regressors in the FMA approach could be an interesting topic for future research.

In a recent paper, Angrist and Pischke (2010) argue that the rise of design-based approaches is the main responsible for the credibility revolution in empirical economics in the last three decades. The treatment effects literature has represented a huge progress in the estimation of more credible causal effects. For instance, randomized experiments and regression discontinuity can be extremely useful for that purpose. Matching estimators might also be very useful, but their identifying exogeneity assumption is conditional on a set of covariates (i.e. it is necessary to control for a group of regressors in order to guarantee the randomness of the treatment assignment). Provided that a set of conditioning (or control) variables is required for the validity of the approach, fragility of results to different conditioning sets can potentially be a cause of concern. Extending the model averaging apparatus to non-parametric matching (or other design-based approaches) might be a fruitful line for future research.

1.8 Appendix of Chapter 1

1.8.1 Asymptotic Theory of FMA Estimators

Suppose the density of the model in Section 1.4 is:

$$f_{true} = f(y, \beta, \gamma) = f(y, \beta_0, \gamma_0 + \delta/\sqrt{N})$$

where β is a parameter present in all models with β_0 its corresponding true value. γ is a vector around its true value γ_0 with perturbation $\delta/\sqrt{N}$. This setting is the local misspecification framework considered in Hjort and Claeskens (2003) for deriving the asymptotic results of FMA estimators. Let $\mu_{true} = \mu(f_{true})$ be the quantity of interest being $\mu(\cdot)$ a known function. The estimator of μ_{true} under the model M_j is given by:

$$\hat{\mu}_{M_j} = \mu(\hat{\beta}_{M_j}, \hat{\gamma}_{M_j}, \gamma_0, M_j^C)$$

where $\hat{\beta}_{M_j}$ and $\hat{\gamma}_{M_j}$ are maximum likelihood estimates and M_j^C is the complement of M_j.

Hjort and Claeskens (2003) analyzed the asymptotic properties of the FMA estimator:

$$\hat{\mu}_{FMA} = \sum_{j=1}^{2^q} \omega_{M_j} \hat{\mu}_{M_j}$$

where ω_{M_j} is a weight function for model M_j given $D_N = \hat{\delta}_{full}$, an estimator of δ under the model with all the q controls (i.e. the full model).

Let us introduce some notation. The score function is given by:

$$\begin{pmatrix} U(y) \\ V(y) \end{pmatrix} = \begin{pmatrix} \partial \log f(y, \beta_0, \gamma_0)/\partial\beta \\ \partial \log f(y, \beta_0, \gamma_0)/\partial\gamma \end{pmatrix}$$

with $(1+q) \times (1+q)$ variance matrix at $(\beta_0, \gamma_0')'$ given by:

$$J = \begin{pmatrix} J_{00} & J_{01} \\ J_{10} & J_{11} \end{pmatrix} \qquad \text{and inverse} \qquad J^{-1} = \begin{pmatrix} J^{00} & J^{01} \\ J^{10} & J^{11} \end{pmatrix}$$

The following theorem corresponds to Theorem 4.1 in Hjort and Claeskens (2003), and it provides the asymptotic distribution of the FMA estimator: If the weight functions ω_{M_j} sum to 1 and have at most a countable number of discontinuities, then:

$$\sqrt{N}\,(\hat{\mu}_{FMA} - \mu_{true}) \xrightarrow{d} \Lambda = \mu_\beta' J_{00}^{-1} \zeta + w'[\delta - \hat{\delta}(D)]$$

where $D \sim N(\delta, \Psi)$ is the limit of D_N, $\Psi = J^{11}$, $\mu_\beta = \frac{\partial\mu}{\partial\beta}$ evaluated at the point (β_0, γ_0'), $\zeta \sim N(0, J_{00})$ independent of D, $\hat{\delta}(D) = \left\{\sum \omega_{M_j} \pi_{M_j}' (\pi_{M_j} \Psi^{-1} \pi_{M_j}')^{-1} \pi_{M_j}\right\} \Psi^{-1} D$, and $w = J_{10} J_{00}^{-1} \mu_\beta - \mu_\gamma$. Finally, let π_{M_j} be the projection matrix mapping δ to δ_j.

1.8.2 Markov Chain Monte Carlo Model Composition

The Markov Chain Monte Carlo Model Composition (MC^3) algorithm proposed by Madigan and York (1995) generates a stochastic process that moves through model space. The idea is to construct a Markov chain of models $M(i), i = 1, 2, ...$ with state space Ψ. If we simulate this Markov chain for $i = 1, ..., N$, then under certain regularity conditions, for any function $h(M_j)$ defined on Ψ, the average:

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

converges with probability 1 to $E\left(h\left(M\right)\right)$ as $N \to \infty$. For example, to compute (1.7) in this fashion, we set $h(M_j) = E(\beta|y, M_j)$.

To construct the Markov chain, we define a neighborhood $nbd(M)$ for each $M \in \Psi$ 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 $\mathbf{q}(M \to M')$. It is the 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.[14]

[14]Koop (2003) is a good reference for the reader interested in developing a deeper understanding of the MC^3 algorithm.

CHAPTER 2

Determinants of Economic Growth: A Bayesian Panel Data Approach

2.1 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 to construct 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, Fernández et al. (2001b) -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 chapter 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 chapter 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 chapter. 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 income 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 chapter is organized as follows. Section 2.2 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 2.3 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 2.4 we briefly describe the data set. The empirical results are presented in Section 2.5. The final section concludes.

[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 income data used (*e.g.* different versions of the Penn World Tables (PWT)).

2.2 Bayesian Model Averaging

A generic representation of the canonical growth regression is:

$$\gamma = \theta X + \varepsilon, \tag{2.1}$$

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 chapter initial GDP will be included with probability 1 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 y -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:

$$g\left(\theta^j|y, M_j\right) = \frac{f\left(y|\theta^j, M_j\right) g\left(\theta^j|M_j\right)}{f\left(y|M_j\right)}, \tag{2.2}$$

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 $P\left(M_j\right)$ we can calculate the posterior model probability using Bayes Rule as:

$$P\left(M_j|y\right) = \frac{f\left(y|M_j\right) P\left(M_j\right)}{f\left(y\right)}. \tag{2.3}$$

Since $P\left(M_j\right)$ does not involve the data, it measures how likely we believe M_j to be the correct model before seeing the data. $f\left(y|M_j\right)$ is often called the marginal (or integrated) likelihood, and is calculated using (2.2) and a few simple manipulations. In particular, if we integrate both sides of (2.2) with respect to θ^j, use the fact that $\int g\left(\theta^j|y, M_j\right) d\theta^j = 1$ (since probability density functions integrate to one), and rearrange, we obtain:

$$f\left(y|M_j\right) = \int f\left(y|\theta^j, M_j\right) g\left(\theta^j|M_j\right) d\theta^j. \tag{2.4}$$

[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

The quantity $f(y|M_j)$ given by equation (2.4) is the marginal probability of the data, because it is obtained by integrating the joint density of (y, θ^j) given y 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:

$$g(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\, g(\theta|y, M_j) \tag{2.5}$$

If one is interested in point estimates of the parameters, one common procedure is to take expectations across (2.5):

$$E(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\, E(\theta|y, M_j)\,. \tag{2.6}$$

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

$$\begin{aligned} V(\theta|y) &= \sum_{j=1}^{2^K} P(M_j|y)\, V(\theta|y, M_j) + \\ &+ \sum_{j=1}^{2^K} P(M_j|y)\left(E(\theta|y, M_j) - E(\theta|y)\right)^2. \end{aligned} \tag{2.7}$$

The posterior variance in (2.7) 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 of that the variable belongs in 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:

$$\text{posterior inclusion probability} = P(\theta_h \neq 0|y) = \sum_{\theta_h \neq 0} P(M_j|y)\,. \tag{2.8}$$

2.2.1 BMA and Growth Regressions

The BMA literature in the growth context (for instance, Fernández et al. (2001b) 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 chapter we extend the use of the BMA methodology to panel data models in the growth context. In subsections 3.2.2 and 3.2.3 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 3.3 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 time-varying regressors and independent of the time-invariant covariates. As we will see in Section 3.3, 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 chapter 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.

[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 chapter 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 3.2.4 for more details).

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 income data used in the estimations. In an attempt to investigate this issue, we replicate our exercises with four different sources of income data: the three last versions of the Penn World Tables (*i.e.* PWT 6.1, PWT 6.2 and PWT 6.3) and the income data reported in the World Development Indicators from The World Bank. We also consider different numbers of candidate regressors in our replications.

2.2.2 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 3.3).

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:

$$y_{it} = \alpha y_{it-\tau} + x_{it}^{\prime j}\beta^j + \eta_i + \zeta_t + v_{it} \quad (t = 1, ..., T)\ (i = 1, ..., N) \tag{2.9}$$

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

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

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 it will be the case in this chapter) 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 (2.9). 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 (2.6) as:

$$E(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\,\widehat{\theta}^j, \tag{2.10}$$

where $\widehat{\theta}^j$ is the WG estimate for θ with the regressor set that defines model j. 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]; therefore:

$$P(M_j|y) = \frac{P(M_j)\,(NT)^{-k^j/2}\,SSE_j^{-(NT)/2}}{\sum_{i=1}^{2^K} P(M_i)\,(NT)^{-k^i/2}\,SSE_i^{-(NT)/2}}, \tag{2.11}$$

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 (2.11) 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 (2.11) 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:

$$W \sim Bin(K, \xi) \tag{2.12a}$$
$$E(W) = K\xi \Rightarrow \xi = \frac{m}{K}. \tag{2.12b}$$

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

[7]For a more detailed discussion of the use of this asymptotic approximation in the BMA context see Section 3.3.2.

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 less 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 chapter.

2.2.3 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 implement the full Bayesian approach with the benchmark priors proposed by Fernández et al. (2001a). 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:

$$y_{it} = \alpha y_{it-\tau} + x_{it}^{\prime j}\beta^j + \phi_1 D_1 + ... + \phi_N D_N + \zeta_t + v_{it} \; (t = 1, ..., T) \; (i = 1, ..., N), \qquad (2.13)$$

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) also hold here, and the error term is supposed to follow a normal distribution. Fernández et al. (2001a) 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:

$$p(\sigma) \propto \sigma^{-1}. \qquad (2.14)$$

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

$$\sigma^2 \left(g_0 Z^{\prime j} Z^j\right)^{-1}, \qquad (2.15)$$

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:

$$B_{ji} = \left(\frac{g_{oj}}{1+g_{oj}}\right)^{\frac{k_j+1}{2}} \left(\frac{g_{oi}+1}{g_{oi}}\right)^{\frac{k_i+1}{2}} \left(\frac{\frac{1}{g_{oi}+1}SSE_i + \frac{g_{oi}}{g_{oi}+1}(y'y)}{\frac{1}{g_{oj}+1}SSE_j + \frac{g_{oj}}{g_{oj}+1}(y'y)}\right)^{\frac{NT}{2}}. \tag{2.16}$$

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:

$$P(M_j) = 2^{-K}. \tag{2.17}$$

The prior in (2.17) 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.

2.2.4 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 – compensates 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:

$$W \sim Bin\,(K, \xi) \tag{2.18}$$

$$\xi \sim Be\,(a, b)\,, \tag{2.19}$$

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:

$$E\,(W) = \frac{a}{a+b}K. \tag{2.20}$$

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 (2.20), 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 can see in Table 2.3, with random ξ the effects of different prior assumptions are much less severe.

2.3 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:

$$E(\theta|y) = \sum\nolimits_{j=1}^{2^K} P(M_j|y)\,\widehat{\theta}^j_{ML}. \tag{2.21}$$

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

The argument behind equation (2.21) 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 (2.21) 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.

2.3.1 The Likelihood Function

The panel data methods employed in Section 3.2 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.

[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.

For a given model M_j we can write:

$$y_{it} = \alpha y_{it-\tau} + x_{it}'^{j}\beta^j + z_i^j\gamma^j + \eta_i + \zeta_t + v_{it} \tag{2.22}$$

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 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 (2.22).

Under assumptions (A3) and (A4) we can write the likelihood as:[10]

$$\begin{aligned} \log f\left(y_i|y_{i0}, x_i^j, z_i^j\right) \propto & -\frac{T-1}{2}\log\sigma_v^2 - \\ & - \frac{1}{2\sigma_v^2}\left(y_i^* - \alpha y_{i(-1)}^* - x_i^{*j}\beta^j\right)'\left(y_i^* - \alpha y_{i(-1)}^* - x_i^{*j}\beta^j\right) - \\ & - \frac{1}{2}\log\omega^2 - \frac{1}{2\omega^2}\left(\overline{y}_i - \alpha\overline{y}_{i(-1)} - \gamma^j z_i^j - \phi^j\overline{x}_i^j - \varphi y_{i0}\right)^2, \end{aligned} \tag{2.23}$$

where $\phi^j = \beta^j + \delta^j$, φ and ω^2 are the linear projection coefficients of $\overline{u}_i$ on $\overline{x}_i^j$, $\overline{u}_i$ is equal to $\eta_i + \overline{v}_i$, and $\overline{v}_i = (1/T)\sum_{t=1}^T v_{it}$. Moreover, y_{i0}, and 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.[11]

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

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

[10]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.

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

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.

2.3.2 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 (2.4). 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)):

$$S = \log f\left(y|\widehat{\theta}_i, M_i\right) - \log f\left(y|\widehat{\theta}_j, M_j\right) - \frac{(k_i - k_j)}{2}\log\left(NT\right), \tag{2.24}$$

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 3.2.2 for a more detailed discussion). As $NT \to \infty$, this quantity, often called the Schwarz criterion, satisfies:

$$\frac{S - \log B_{ij}}{\log B_{ij}} \to 0 \tag{2.25}$$

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

$$BIC = -2S = -2\log B_{ij}. \tag{2.26}$$

Despite 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[12] the error is $O(n^{-1/2})$ instead of $O(1)$. This error is much smaller and it does tend to zero as the sample size increases.

[12] 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.

The value of BIC for model M_j denoted BIC_j, is the approximation to $2logB_{0j}$ given by (2.26), 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 (2.3) as:

$$\begin{aligned} P(M_j|y) &= \frac{f(y|M_j)P(M_j)}{\sum_{i=1}^{2K} f(y|M_i)P(M_i)} = \qquad (2.27) \\ &= \frac{\frac{f(y|M_j)}{f(y|M_h)} f(y|M_h)P(M_j)}{\sum_{i=1}^{2K} \frac{f(y|M_i)}{f(y|M_h)} f(y|M_h)P(M_i)} = \\ &= \frac{B_{jh} f(y|M_h)P(M_j)}{\sum_{i=1}^{2K} B_{ih} f(y|M_h)P(M_i)} = \\ &= \frac{B_{j0}P(M_j)}{\sum_{i=1}^{2K} B_{i0}P(M_i)}, \end{aligned}$$

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

Given the above, instead of integrating to obtain the marginal likelihood in (2.4), we will use the following result:

$$f(y|M_j) \propto \exp\left(-\frac{1}{2}BIC_j\right), \qquad (2.28)$$

and therefore:

$$P(M_j|y) = \frac{P(M_j)\exp\left(-\frac{1}{2}BIC_j\right)}{\sum_{i=1}^{2K} P(M_i)\exp\left(-\frac{1}{2}BIC_i\right)}. \qquad (2.29)$$

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

$$\frac{P(M_i|y)}{P(M_j|y)} = \frac{P(M_i)}{P(M_j)} \frac{\exp\left(-\frac{1}{2}BIC_i\right)}{\exp\left(-\frac{1}{2}BIC_j\right)}. \qquad (2.30)$$

2.3.3 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.

2.4 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 chapter.[13] 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.

2.4.1 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 income, 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

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

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, since this measure is sometimes criticized because it only captures the volume of trade and not 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 (1997). It is worth mentioning that the SW indicator has its own limitations as pointed out by Rodríguez and Rodrik (2000). The final objective in this chapter is to conclude which measure of openness is a better (in the sense of more robust) proxy.

- 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 chapter 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 the Freedom House, and they are freely available at http://www.freedomhouse.org. Barro and Lee (1994b) 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 access (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 minimal 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 (1994b) and Gallup et al. (2001) respectively. The geographical variables and fixed factors considered in this chapter 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.

2.5 Results

2.5.1 Panel BACE-SDM and Panel BMA-FLS Results

Table 2.3 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's 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 2.3, 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 2.4 shows the posterior inclusion probability, the posterior mean and the posterior standard error for 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. 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 Fernández et al. (2001b). This is an indication that country specific effects matter and make a difference in this respect. Note however, that given the nature of this chapter, the main conclusions will be obtained according to the results presented in the next subsection (Table 2.5).

2.5.2 BAMLE Results

Results when applying the BAMLE approach with PWT 6.2 income data for the whole period are summarized in Table 2.5. 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. 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 share can be considered as robust determinants of economic growth. Despite 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, despite it is 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 error is bigger than its posterior mean. This means that this variable is associated with economic growth, but we can not 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.[14] 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 2.1. 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 (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 2.5 implies a rate of conditional convergence of

[14] I thank an anonymous referee for pointing this out.

$\lambda = 0.006$. This suggests that once we control for model uncertainty and other potential inconsistencies (*e.g.* omitted variable and endogeneity biases), the point estimate of the convergence rate is similar to the standard finding in cross-sectional studies. However, in our case, we conclude that there is no evidence of conditional convergence because the posterior standard error is larger than the posterior mean for the convergence parameter.

2.5.3 Sensitivity Results

Ciccone and Jarocinski (2009) point out that agnostic BMA approaches lead to conclusions that are sensitive across available sources of international income data. They compare the different Posterior Inclusion Probabilities (PIP) emerging when using alternative sources of data on income. For each variable j they estimate its PIP using PWT 6.1 and PWT 6.2 income 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 this differences are substantial.

In an attempt to further investigate this issue, Table 2.6 presents measures of sensitivity of the results when using different sources of international income data. We compare the results obtained with the baseline income data used in this chapter (PWT 6.2) with the other two available versions of the Penn World Tables (PWT 6.1 and PWT 6.3) and the income data reported in the World Development Indicators 2005 from The World Bank. More concretely, Table 2.6 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.[15] 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 2.6 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 Tables seems to produce more stable results than previous revisions. On the other hand, results using WDI 2005 and PWT 6.2 income data are more sensitive than across different versions of the PWT.

2.6 Concluding Remarks

In spite of a huge amount of empirical research, the drivers of economic growth are not well understood. This chapter attempts to provide insights on the growth puzzle by extending the

[15] 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 chapter (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 number of available observations for estimation.

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 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 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.[16]

As a final remark, it is worth mentioning that the dynamic panel estimator proposed in this chapter 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.

[16]From a frequentist point of view this means that the convergence rate is not significantly different from zero.

2.7 Appendix of Chapter 2

2.7.1 Computational Appendix

For the implementation of the empirical approaches described in the chapter, 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 $3.4\text{x}10^{10}$. 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 $\mathbf{q}(M \to M')$. It is the 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.[17]

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

[17] Koop (2003) is a good reference for the reader interested in developing a deeper understanding of the MC^3 algorithm.

2.7.2 Data Appendix

Table 2.1: 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 minimal distance (km) from New York, Rotterdam, or Tokyo
Tropical Area	Gallup et al.	Fraction of land area in geographical tropics

Table - 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 opennes 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

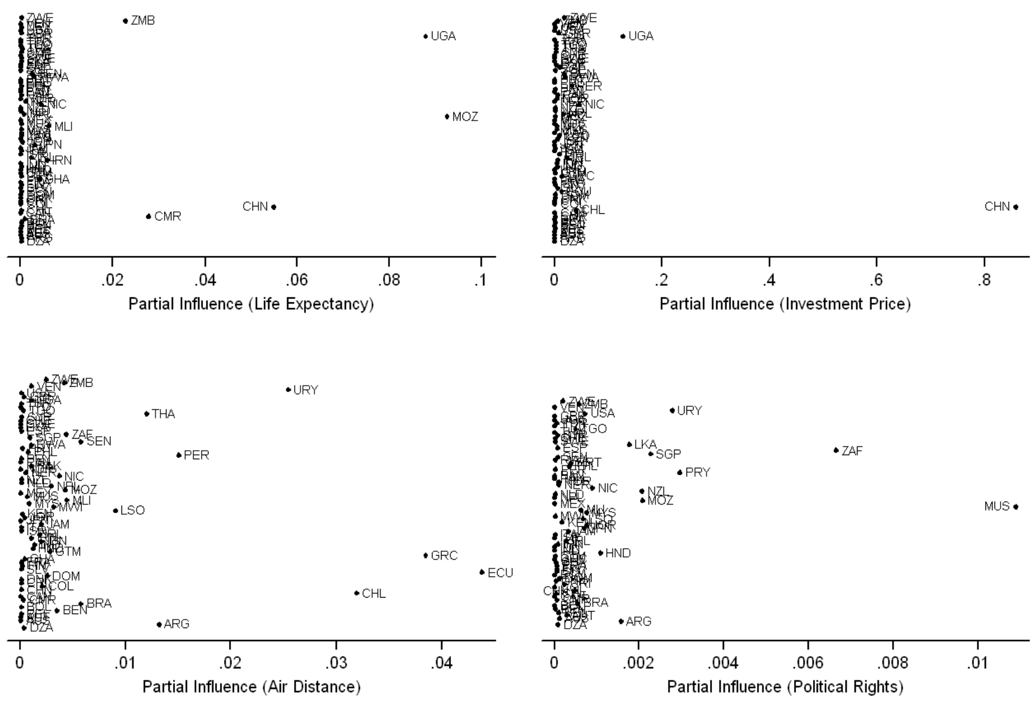
PWT 6.2 refers to Penn World Tables version 6.2 and it can be fount at http://pwt.econ.upenn.edu. WDI 2005 refers to World Development Indicators 2005. Data from Barro and Lee, Gallup et al., and Sachs and Warner is available at http://www.cid.harvard.edu/ciddata/ciddata.html. Finally, data from Freedom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

Table 2.2: List of Countries in Chapter 2

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	

2.7.3 Figures

Figure 2.1: 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".

2.7.4 Tables

Table 2.3: 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 2.4: Panel SDM-FLS Approaches Results

Variable	Posterior Inclusion Probability		Posterior Mean		Posterior Standard Error	
	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 2.5: BAMLE Approach Results

Variable	Posterior Inclusion Probability	Posterior Mean	Posterior Standard Error
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 3.3. All results in this Table are based on prior assumptions $m = 5$ and ξ Random.

Table 2.6: 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 income 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.

CHAPTER 3

Panel Growth Regressions with General Predetermined Variables: Likelihood-Based Estimation and Bayesian Averaging

3.1 Introduction

Due to model uncertainty, Bayesian Model Averaging —henceforth BMA— methods applied to growth regressions have recently been incorporated into the toolkit of empirical growth researchers. However, the recent literature on BMA and growth is always based on the questionable assumption of strict exogeneity of all growth determinants.[1] This is the first article that presents estimates of causal effects simultaneously considering the issues of endogeneity and model uncertainty in the growth context. In order to apply the BMA methodology, or the Bayesian apparatus in general, we need suitable likelihoods. For this purpose, this chapter introduces a likelihood function for a dynamic panel data model with general endogenous variables and fixed effects. The key and new ingredient of this likelihood is the specification of an unrestricted feedback process for the regressors since they are not treated as exogenous. Moreover, the resulting likelihood-based estimator is shown to outperform standard GMM alternatives in finite samples.

As pointed out by Durlauf et al. (2005), the stylized facts of economic growth have led to two major issues in the development of formal econometric analyses of growth. The first one revolves around the question of convergence: are contemporary differences in growth rates across countries transient over sufficiently long time horizons? The second issue concerns the identification of growth determinants: which factors seem to explain observed differences in aggregate economies? These two questions have been addressed by a huge literature on empirical growth regressions. However, this industry is plagued by econometric inconsistencies that arise not only when estimating an empirical growth model (i.e. endogeneity of growth determinants) but also when selecting that model (i.e. model uncertainty). In this chapter I argue that once these issues are accounted for, the empirical results are in contrast to previous consensus in the literature: on the one hand, the estimated convergence rate is not significantly different from zero; on the other hand, there is only one variable, the investment ratio, that robustly causes economic growth.

The issue of model uncertainty emerges because theory does not provide enough guidance for selecting the proper empirical model. Model averaging techniques construct parameter estimates that formally address the dependence of model-specific estimates on a given model. Even though there are many papers that apply BMA techniques to the growth context (e.g. Fernández et al. (2001b) and Sala-i-Martin et al. (2004)), they are all founded on the problematic exogeneity assumption of the growth determinants. Intuitively, these papers estimate millions of models in order to address model uncertainty, but the estimation of all these models is based on the exogeneity assumption which is very probably violated in the growth context. Having said that, it is true that there seems to be consensus on BMA as the most promising solution to model uncertainty.

The endogeneity issue is still unsolved in the growth framework. Problems with estimating an empirical growth model are well known. The right-hand side variables are typically endogenous and measured with error. Omitted variable bias also arises because of the pres-

[1] From a time series perspective, a similar situation is also present in the BMA forecasting literature where the predictors are assumed to be strictly exogenous (see Stock and Watson (2006), page 541)

ence of unobservable time-invariant country-specific characteristics correlated with one or more regressors. The most prominent way to address these problems is the use of panel data econometric techniques that allow for country-specific fixed effects in the empirical model.[2] In particular, first-differenced GMM estimators applied to dynamic panel data models has been the most promising econometric method in empirical growth research. This estimation procedure addresses the question of correlated individual effects and the issue of endogeneity and it was first proposed in the econometrics literature by Holtz-Eakin et al. (1988) and Arellano and Bond (1991), while in the growth context it was first considered by Caselli et al. (1996).

Despite its important advantages over simple cross-section regressions and other estimation methods for dynamic panel data models, it is now well known that in the growth context this method suffers from large finite-sample biases. Given the variables considered in empirical growth models, the time series are persistent and the number of observations in the cross-section dimension is typically small. Under these conditions, the first-differenced GMM estimator is poorly behaved because lagged levels of the variables are only weak instruments for subsequent first-differences. This weak-instruments problem may be present in other situations with highly persistent data in a small-T panel setting.

By assuming mean stationarity of the variables, we can exploit additional moment conditions and employ the so-called system-GMM estimator as proposed in Arellano and Bover (1995) in order to alleviate the described weak-instruments problem. However, in the analysis of country panel data, Barro and Sala-i-Martin (2003) described some examples —like data sets that start at the end of a war or other major historical event— in which one would not expect initial conditions to be distributed according to the steady state distribution of the process in any dimension. Therefore, if we are willing to avoid stationarity assumptions, as we are in general, and specially in the growth context, there is no better alternative proposed for this situation. To overcome this issue, this chapter presents a feasible likelihood-based estimator in a panel data context which is asymptotically[3] equivalent to one-step first-differenced GMM augmented with moments implied by the serial correlation properties of errors.[4] This maximum likelihood estimator alleviates the weak-instruments problem in finite samples without resorting to auxiliary stationarity assumptions.

I also argue that the estimator can be applied to a broad range of situations in addition to growth regressions. One prominent example is the estimation of production functions in which we typically face two problems: (i) the regressors (employment and stock of capital) are potentially correlated with firm-specific fixed effects and productivity shocks, and, (ii) both employment and capital are highly persistent processes. Not surprisingly, first-differenced GMM has poor finite-sample properties in this context. Some authors have proposed to incorporate stationarity assumptions to the model and employ the denominated system-

[2]Typical growth panels are based on a sample of N countries observed over ten or five-year periods. Despite some exercises are carried out with five-year periods, current data availability allows me to focus on ten-year periods in order to avoid business cycle effects, following Barro and Sala-i-Martin (2003).

[3]I refer here to fixed-T and $N \to \infty$ asymptotics.

[4]The additional moments are quadratic restrictions of the type discussed in Ahn and Schmidt (1995).

GMM estimator in order to alleviate the weak-instruments problem (see for example Blundell and Bond (2000)). Again, as in the growth context, the likelihood-based estimator proposed in this chapter is able to solve the weak-instruments problem present in the estimation of production functions without making any additional assumption. By the same token, there are many other situations in which the econometric issues just described are also present.

In the single equation case, it is well documented in the literature that the effect of weak-instruments on the distribution of two-stage least squares (2SLS) and limited information maximum likelihood (LIML) differs substantially in finite samples despite the fact that both estimators have the same asymptotic distribution. Although the distribution of LIML is centered at the parameter value, 2SLS is biased toward ordinary least squares (OLS). On the other hand, since LIML has no finite moments regardless of the sample size, its distribution has thicker tails than that of 2SLS. In terms of numerical comparisons of median bias, interquartile ranges, and rates of approach to normality, Anderson et al. (1982) concluded that LIML was to be strongly preferred to 2SLS, particularly if the number of instruments is large.

In the panel setting considered in this chapter, the number of instruments increases with the time series dimension (T), and, therefore, the model generates many overidentifying restrictions even for moderate values of T, although the quality of these instruments is often poor. In order to construct the likelihood function, there are T structural equations, but how to complete the model with the reduced form equations is not straightforward[5]. Two different possibilities are presented in this chapter. After concentrating the resultant likelihood function, the maximum likelihood estimator (i.e. the LIML counterpart of GMM estimators of panel data models with general endogenous or predetermined variables and fixed effects) is easy to apply by means of numerical optimization methods.

The finite-sample behavior of the sub-system LIML estimator developed in this chapter is investigated via Monte Carlo simulations in an experimental design closely calibrated to panel cross-country growth regressions. The Monte Carlo results show that sub-system LIML has negligible biases in contrast to the Arellano-Bond GMM estimator, which has large biases in most of the cases I consider. Therefore, the main conclusion is that the likelihood-based estimator I propose in this chapter is strongly preferred to standard GMM estimators in terms of finite-sample performance.

Regarding the empirical growth literature, closely related to the econometric issues mentioned above, there is the convergence debate. After two decades of research the question is still unanswered: Is there conditional convergence across countries? Some authors consider that the available empirical evidence supports the conditional convergence hypothesis predicted by the neoclassical growth model. However, from a skeptical point of view, the lack of reliable estimates of the convergence parameter in growth regressions is enough to hamper consensus on the answer of this relevant question. Furthermore, despite some progress has been done, there is no clear evidence on the most prominent variables in fostering economic

[5]In the pure autorregresive case Alvarez and Arellano (2003) among others have derived the likelihood function. To the best of my knowledge, thus far there is no paper deriving the likelihood for the case with general predetermined variables

growth. The reason is that all previous studies attempting to solve this issue are based on partial correlations and not causal effects (e.g. Sala-i-Martin et al. (2004) and Fernández et al. (2001b)).

Given the above, after considering all potential sources of biases and inconsistencies (i.e. after combining the BMA methodology with the proposed likelihood-based estimator), I obtain two results that are in contrast to previous consensus in the literature. On the one hand, I find that conditional convergence is not present across the countries in my sample. In particular, the estimated speed of convergence is 0.73%, but it is not significantly different from zero. This result would lead us to conclude that the hypothesis of no conditional convergence cannot be rejected given the available data. On the other hand, I conclude that there is only one variable that seems to cause economic growth, the investment ratio. However, I obtain further evidence that allows me to conclude that some variables, such as population or life expectancy, in spite of having a statistically insignificant effect on growth, should be included as controls in growth regressions. This is so, because the models that include these variables are the best models in fitting the data.

The remainder of the chapter is organized as follows. Section 3.2 describes the construction of the likelihood function in the context of a dynamic panel data model with feedback. Monte Carlo evidence on the finite-sample behavior of the estimator is provided in Section 3.3. In Section 3.4 I estimate some different specifications of empirical growth models with the proposed estimator. Results from combining the estimator and model averaging techniques are presented in Section 3.5. Finally, Section 3.6 concludes and auxiliary results are gathered in the Appendix.

3.2 Dynamic Panel Data with Feedback: Likelihood-Based Estimation

Consider the following panel data model:

$$y_{it} = \alpha y_{it-1} + x_{it}'\beta + w_i'\delta + \eta_i + \zeta_t + v_{it} \tag{3.1}$$

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, w_i, \eta_i\right) = 0 \qquad (t = 1, ..., T)(i = 1, ..., N) \tag{3.2}$$

where x_{it} and w_i are vectors of variables of orders k and m respectively, and x_i^t denotes a vector of observations of x accumulated up to t: $x_i^t = (x_{i1}', \ldots, x_{it}')'$

The predetermined nature of the lagged dependent variable is considered in assumption (3.2). The model also relaxes the strict exogeneity assumption for the x variables that are also considered as predetermined (this is why we can refer to the model as having general predetermined variables). In particular, the assumption in (3.2) allows for feedback from lagged values of y to the current value for x. Moreover, it implies lack of autocorrelation in v_{it} since lagged vs are linear combinations of the variables in the conditioning set. A notational remark is that the model is written in such a way that the initial observation for

y is y_{i0} and for the xs the initial observation is x_{i1}. Both are observed and, in any case this is just a matter of notation.

I also include m strictly exogenous regressors that may or may not have temporal variation. In the remaining of the exposition I assume that all the w variables have no variation within time. While allowing for time varying strictly exogenous w variables is straightforward in this context, in the spirit of Hausman and Taylor (1981) I prefer to stress the possibility of identifying the effect of time-invariant variables in addition to the unobservable time-invariant fixed effect. This is possible by assuming lack of correlation between the w variables and the unobservable fixed effects η_i.

Note that in addition to the individual specific fixed effects η_i, I also include the term ζ_t in (3.1), that is, time dummies are present in the model in order to capture unobserved common factors across units in the panel and, therefore, I allow for these particular forms of cross-sectional dependence. In practice, this is done by simply working with cross-sectional de-meaned data. In the remaining of the exposition, I assume that all the variables are in deviations from their cross-sectional mean.

Models like the one presented in equations (3.1)-(3.2) are typically estimated by first-differenced generalized method of moments. However, the conclusion from a sizeable Monte Carlo literature on the finite-sample properties of this GMM estimators is that they can be severely biased when weak instruments (persistent series) are present (e.g. Arellano and Bond (1991), Blundell and Bond (1998) and Alonso-Borrego and Arellano (1999) amongst others). In order to alleviate this problem, some alternatives have been proposed in the literature (see for example Hansen et al. (1996) and Alonso-Borrego and Arellano (1999)). On the other hand, given the available evidence in the single equation case, likelihood-based estimators are also good candidates in the face of the weak-instruments problem in this setting. Moreover, the availability of a proper likelihood function would allow us to combine the apparatus of likelihood-based inference and the Bayesian framework with dynamic panel data models with general predetermined variables and fixed effects.

Previous likelihood-based approaches in dynamic panel data models only consider the case of strictly exogenous regressors (see for example Bhargava and Sargan (1983)). Therefore, the focus was on the distribution of y_i^T conditional on the regressors and, sometimes on the initial observation y_{i0}. Moreover, it is possible to either condition on the fixed effect η_i or work with the distribution marginal on the effects (see Arellano (2003) for more details). In any case, the distribution of the regressors is not specified since they are considered as strictly exogenous. If this assumption is not true, as it is the case in many applications such as growth regressions or macro forecasting applications, the likelihood will be fundamentally misspecified. Here instead I present the likelihood function for dynamic panel data models with general predetermined variables and fixed effects.

3.2.1 Completing the Model with an Unrestricted Feedback Process

In contrast to a model with only strictly exogenous explanatory variables, the specification of the model with predetermined variables is incomplete in the sense that in itself it does not lead to a likelihood once we add an error distributional assumption. To complete the model in a way that is not restrictive, I specify the feedback process as a linear projection of the non-exogenous variables on all available lags, having period-specific coefficients. The complete model is therefore as follows:

$$y_{i0} = w_i'\delta_y + c_y\eta_i + v_{i0} \tag{3.3a}$$
$$x_{i1} = \Delta_1 w_i + \gamma_{10}y_{i0} + c_1\eta_i + u_{i1} \tag{3.3b}$$
$$y_{i1} = \alpha y_{i0} + x_{i1}'\beta + w_i'\delta + \eta_i + v_{i1} \tag{3.3c}$$

and for $t = 2, ..., T$:

$$x_{it} = \Delta_t w_i + \gamma_{t0}y_{i0} + ... + \gamma_{t,t-1}y_{i,t-1} + \Lambda_{t1}x_{i1} + ... + \Lambda_{t,t-1}x_{i,t-1} + c_t\eta_i + u_{it} \tag{3.3d}$$
$$y_{it} = \alpha y_{i,t-1} + x_{it}'\beta + w_i'\delta + \eta_i + v_{it} \tag{3.3e}$$

> **Remark:** Note that by writing the system as in (3.3a)-(3.3e) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$, since otherwise I should have added the equation $\eta_i = w_i'\delta_\eta + e_i$ in order to complete the system. Therefore, assuming that $\delta_\eta = 0$ is enough to guarantee identification of δ in (3.1).

This is a system of $T(k+1)+1$ equations where δ_y and c_t are vectors of parameters of order m and k respectively, c_y is a scalar, and γ_{th} is the $k \times 1$ vector:

$$\gamma_{th} = (\gamma_{th}^1, \ldots, \gamma_{th}^k)' \quad (t = 1, \ldots, T) \;\; (h = 0, \ldots, T-1)$$

Moreover, Δ_t and Λ_{th} are matrices of parameters of orders $k \times m$ and $k \times k$, respectively, and u_{it} is a $k \times 1$ vector of prediction errors.

On the other hand, I also define the $T(k+1)+2$ column vector of errors:

$$\Xi_i = (\eta_i, v_{i0}, u_{i1}', v_{i1}, \ldots, u_{iT}', v_{iT})'$$

and the $T(k+1)+1 \times 1$ vector of data for individual i:

$$R_i = (y_{i0}, x_{i1}, y_{i1}, \ldots, x_{iT}, y_{iT})'$$

Finally, in order to rewrite the system in matrix form, I define the $T(k+1)+1 \times T(k+1)+1$ lower triangular matrix of coefficients B as:

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\gamma_{10} & I_k & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\alpha & -\beta' & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\gamma_{20} & -\Lambda_{21} & -\gamma_{21} & I_k & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\alpha & -\beta' & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & 0 & 0 & 0 \\ -\gamma_{T0} & -\Lambda_{T1} & -\gamma_{T1} & -\Lambda_{T2} & -\gamma_{T2} & \dots & -\gamma_{T,T-1} & I_k & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & -\alpha & -\beta' & 1 \end{pmatrix}$$

And the matrices D and C of orders $T(k+1)+1 \times T(k+1)+2$ and $T(k+1)+1 \times m$ respectively:

$$D = \begin{pmatrix} c_y & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ c_1 & 0 & I_k & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ c_2 & 0 & 0 & 0 & I_k & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_T & 0 & 0 & 0 & 0 & 0 & I_k & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad C = \begin{pmatrix} \delta'_y \\ \Delta_1 \\ \delta' \\ \vdots \\ \Delta_T \\ \delta' \end{pmatrix}$$

Given the above, I am now able to write the system in matrix form as follows:

$$BR_i = Cw_i + D\Xi_i$$

where:

$$Var\left(\Xi_i\right) = \Omega = \begin{pmatrix} \sigma_\eta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{v_0} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma_{u_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{v_1} & 0 & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & 0 & \Sigma_{u_T} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{v_T} \end{pmatrix}_{T(k+1)+2 \times T(k+1)+2}$$

and Σ_{u_t} is a $k \times k$ matrix.

This parametrization of the complete model is labeled as Full Covariance Structure (FCS) representation. Moreover, under normal errors the log-likelihood of the model can be written as:

$$L = -\frac{N}{2}\ln\det\left(B^{-1}D\Omega D'B'^{-1}\right) - \frac{1}{2}tr\left\{\left(B^{-1}D\Omega D'B'^{-1}\right)^{-1}\left[R-W(B^{-1}C)'\right]'\left[R-W(B^{-1}C)'\right]\right\} \quad (3.4)$$

where R and X_t are the following matrices:

$$R = \begin{pmatrix} Y_0 & X_1 & Y_1 & \ldots & X_T & Y_T \end{pmatrix}_{N\times T(k+1)+1}$$
$$X_t = \left(X_t^1, \ldots, X_t^k\right)_{NXk}$$

and W is the $N \times m$ matrix $W = (w_1, w_2, \ldots, w_N)'$.

It is important to remark here that the maximizer of L is a consistent and asymptotically normal estimator regardless of non-normality. More specifically, the resultant first order conditions correspond to a GMM problem with a convenient choice of weighting matrix (see Arellano (2003) pp.71-73).

Note also that the coefficients matrix B includes γ_{th} and Λ_{th} that are the vector and matrix that gather all the feedback process from lagged ys to current xs and the dynamic relationships between the x variables respectively. The parameters corresponding to the dynamic relationships between the xs are not of central interest for our model, but in principle, they also need to be estimated. In practice this might be a concern since the number of them is enormous.

On the other hand, the variance-covariance matrix of the errors Ω is block-diagonal. An interesting feature of this model is that there is a one-to-one mapping between the parameters in B and the elements of Ω. More specifically, any coefficient in γ_{th} or Λ_{th} restricted to be zero in B will automatically be translated into an additional non-zero element in Ω in order to satisfy the same number of restrictions imposed by the model. Further developing this feature, I present in the Appendix another parametrization (labeled as Simultaneous Equation Model (SEM) representation) that captures the feedback process and the dynamic relationships between the xs in the variance-covariance matrix of the system. This SEM parametrization turns out to be useful in practice because it allows me to concentrate out all the parameters of the dynamic relationships between the xs. This concentration, described in Appendix, drastically reduces the number of paramaters to be estimated.

3.3 Monte Carlo Simulation

In this section, I provide some Monte Carlo evidence on the finite-sample behavior of the likelihood-based estimator proposed in the previous section. The purpose is to study its finite-sample properties in relation to the commonly used first-differenced GMM and Within-Group estimators.

3.3.1 Model and Estimators

Let us consider a dynamic panel data model with feedback and fixed effects as follows:

$$y_{it} = \alpha y_{it-1} + \beta_1 x^1_{it-1} + \beta_2 x^2_{it-1} + \eta_i + v_{it} \tag{3.5}$$

$$E\left(v_{it} \mid y_{it-1}, ..., y_{i0}, x^1_{it-1}, ..., x^1_{i0}, x^2_{it-1}, ..., x^2_{i0}, \eta_i\right) = 0 \tag{3.6}$$

Suppose we have a random sample of individual time series of size T: $(w'_{i1}, ..., w'_{iT})'$ where $w_{it} = (y_{it}, x^1_{it}, x^2_{it})'$ and $(i = 1, ..., N)$. On the other hand, I assume that initial observations $w_{i0} = (y_{i0}, x^1_{i0}, x^2_{i0})'$ are observed. I further assume that the initial observations and the fixed effect are jointly normally distributed[6] with unrestricted mean vector and covariance matrix. In other words: (i) feedback is allowed from lagged y to current x's. (ii) Stationarity assumptions of any type are avoided. (iii) Individual fixed effects correlated with the regressors are included.

The Monte Carlo design tries to mimic as close as possible the Solow model environment. For this purpose, parameter values are fixed according to the results obtained in the estimation of a VAR process for the variables GDP (y), investment ratio (x^1) and population growth (x^2) over the period 1960-2000. Using these parameter estimates I simulate random samples according to a structural VAR data generating process. Specifically, the employed parameter values correspond to the estimates obtained when estimating the VAR process using ten-year periods data, the baseline specification in this chapter. On the other hand, since five-year periods are also commonly considered in empirical panel growth regressions, for the purpose of robustness, I also conduct a set of Monte Carlo simulations using parameter values calibrated to five-year periods data. These additional results and more details on the Monte Carlo design can be found in the Appendix.

Three alternative estimators are applied to the simulated samples. I first consider the Within-Group (WG) estimator of $(\alpha, \beta_1, \beta_2)'$. This is given by the slope coefficients in an OLS regression of y on lagged w and a full set of individual dummy variables, or equivalently by the OLS estimate in deviations from time means or orthogonal deviations. Assumptions required for consistency of the WG estimator (i.e. strict exogeneity of the regressors) are not satisfied in our setting. However WG is considered in order to make comparisons with first-differenced GMM (diff-GMM) since similarities between both are typically considered as indication of the presence of weak instruments in the diff-GMM estimates (see Bond et al. (2001)).

Secondly, I consider the diff-GMM estimator commonly employed in panel growth regressions since Caselli et al. (1996). The assumption in equation (3.6) implies a set of linear moment conditions of the form:

$$E[w_i^{t-2}(\Delta y_{it} - \alpha \Delta y_{it-1} - \beta_1 \Delta x^1_{it-1} - \beta_2 \Delta x^2_{it-1})] = 0 \tag{3.7}$$

[6] Note that the consistency of the estimators I consider in the Monte Carlo exercise is unaffected by the normality assumption (see Arellano (2003) pp.71-73).

In our case, this moment conditions are exploited using the optimal one-step GMM estimator under "classical" errors and it is labeled as diff-GMM. This estimator is consistent under the same assumptions as the likelihood-based estimator proposed in this chapter. Given the persistence of the series considered in the growth context, the diff-GMM estimator is expected to suffer from weak instruments in finite samples.

The maximum likelihood estimator proposed in the previous section is expected to alleviate the weak-instruments problem in finite samples. Therefore it is also considered in our experiment in order to study its finite-sample performance in relation to diff-GMM. This estimator is labeled as sub-sys LIML since it can be interpreted as a sub-system LIML estimator.

Under homoskedasticity, sub-system LIML is asymptotically equivalent to a GMM estimator that in addition to (3.7) uses the following moments implied by lack of serial correlation:

$$E[\Delta v_{i,t-1} u_{it}] = 0 \qquad\qquad (t = 3, ..., T)$$

where $u_{it} = \eta_i + v_{it}$. Thus, in the comparison between sub-system LIML and diff-GMM there are two sources for different performance. First, the extra moments and second the finite-sample differences.

3.3.2 Results

Table 3.4 reports sample medians, percentage median bias, interquartile ranges, and median absolute errors (MAE's) for WG, diff-GMM and sub-sys LIML estimators for the model in equations (3.5)-(3.6) (means and standard deviations are not reported because the sub-system LIML estimators can be expected to have infinite moments).

In the baseline specification in Panel A, N is fixed to 100 since it is the number of cross-section observations we find in a typical growth regression. On the other hand, given the main focus of this chapter is on ten-year periods over the years 1960-2000, $T = 4$ is the number of available time series observations. In this baseline experiment, which replicates as close as possible the situation in empirical panel growth regressions, sub-sys LIML clearly outperforms diff-GMM. In terms of median bias, diff-GMM is badly biased in all the three coefficients while sub-system LIML has always much smaller biases that are almost negligible in the cases of α and β_2. Note here that the percentage of median bias is not informative when comparing estimates across different coefficients since it depends on the magnitude of the true coefficient. However it is illustrative for comparisons between different estimates of the same coefficient. For example, the percentage of bias in α for sub-system LIML is only 5.2% while for WG and diff-GMM this percentage is huge, 55.2% and 53.7% respectively. An additional remark, is that diff-GMM estimates are more similar to WG estimates than to the true values in the case of the autorregresive parameter, and this is an indication of weak instruments in the diff-GMM estimator. On the other hand, looking at the interquartile range (iqr), WG has always less dispersion than diff-GMM and sub-sys LIML as expected. However, the dispersion of sub-sytem LIML is very similar to that of diff-GMM and even smaller for the α parameter. This means that the higher probability of outliers in LIML estimators is not a big concern in this particular application. Finally, attending to MAE's, sub-sys LIML always performs clearly better than diff-GMM. MAE summarizes information on the performance of the estimator in terms of both bias and dispersion. Summing up, the conclusion from Panel A in Table 3.4 is that sub-system LIML clearly ourtperforms diff-GMM in the typical situation that an empirical growth researcher faces when using ten-year periods over the post-war sample 1960-2000.

In Panels B and C of Table 3.4, the results with $N = 500$ and $N = 1000$ are presented for illustrating the performance of the estimators in larger samples. In principle this is not a realistic situation in the cross-country growth context since there are not so many countries in the world. However, one could use regional data and have a sample size of a magnitude similar to 500 in the cross-section dimension. In any case, the purpose of this experiment is to investigate the relative performance of diff-GMM and sub-sys LIML in larger samples (larger in the cross-section dimension) since both estimators are consistent as $N \to \infty$ and T remains fixed. The performance of WG is not affected by increasing N since the WG bias comes from the small sample size in the time series dimension. Therefore, in terms of median bias, the WG results are practically the same in Panels A, B, and C. However, as expected, diff-GMM performance substantially improves as N increases in terms of median bias and dispersion. This improvement is not so substantial for sub-sys LIML since its performance is already reasonably satisfactory with $N = 100$ as shown in Panel A. However, looking at MAE's as a summary measure, sub-system LIML is still considerably better than diff-GMM in all cases. In any event, while sub-sys LIML biases become insignificant for moderate values of N, the diff-GMM biases are not negligible even with $N = 1000$. This would lead us to the conclusion that, with four time series observations, in order to consider the consistency results valid in this application, diff-GMM requires sample sizes larger than 1000 in the cross-section dimension, which seems clearly implausible in the growth context.

Three additional experiments based on $T = 8$ are presented in the three bottom panels of Table 3.4. I also consider these experiments because five-year periods are commonly considered in the panel growth literature, and, if we consider the post-war period 1960-2000, we would end up with eight time series observations. Panels D, E, and F present the results with $N = 100$, $N = 500$, and $N = 1000$ respectively. These results confirm the patterns previously described (i.e. sub-sys LIML clearly outperforms diff-GMM for all sample sizes in the cross-section dimension) but now, with $T = 8$, the biases and interquartile ranges for both diff-GMM and sub-sys LIML are always smaller for a given value of N. This means that the performance of both estimators clearly improves as the number of time series observations increases. As expected, this is also true in the case of WG.

Finally, all the experiments previously described are conducted again but using different parameter values for the purpose of robustness. Both the employed parameter values and the results are available in the Appendix. These additional results confirm the patterns that emerge from Table 3.4. Given the above, the main conlusion from our Monte Carlo study is that, in the growth context, the likelihood-based estimator (sub-sys LIML) presented in this chapter clearly outperforms the commonly used diff-GMM estimators in finite samples. This is true even when the number of available cross-section observations is around 1000.

3.4 Empirical Growth Regressions

The neoclassical framework is the basis for most empirical growth research. Departing from a generic one-sector growth model, in either its Solow-Swan or Ramsey-Cass-Koopmans variant, it is usual to assume that aggregate output obeys a Cobb-Douglas production function and then obtain a canonical cross-country growth regression of the form:

$$\gamma_i = \beta \ln y_{i0} + \psi X_i + \epsilon_i \tag{3.8}$$

where $\gamma_i = t^{-1}(\ln y_{it} - \ln y_{i0})$ represents the growth rate of output per worker between 0 and t. On the other hand, X_i is a vector of variables that represents not only the growth determinants suggested by the the Solow-Swan growth model but also additional determinants that allow for predictable heterogeneity in the steady state. These regressions are sometimes called Barro regressions, given Barro's extensive use of such regressions to study alternative growth determinants starting with Barro (1991). These kind of regressions have been widely used trying to address two major themes in the formal empirical analysis of growth: the identification of growth determinants and the question of convergence.

As previously stated, most of the growth econometrics literature is based on equation (3.8). An important objective of the present chapter is to solve the problems that are still present in these empirical growth regressions from an econometric perspective. In particular, I address the issues of endogeneity, omitted variables, model uncertainty, measurement error, and, to some extent, parameter heterogeneity. By doing so, I will then be able to shed some light on the two issues mentioned above.

There is an important variant of the baseline empirical growth regression in (3.8) that can be called the canonical panel growth regression:

$$\ln y_{i,t} = (1+\beta)\ln y_{i,t-1} + \psi X_{i,t-1} + \eta_i + \zeta_t + v_{i,t} \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{3.9}$$

where η_i is a country-specific fixed effect that allows considering unobservable heterogeneity across countries (since this term is country specific, we can interpret it as allowing for some kind of parameter heterogeneity across countries), and ζ_t is a period-specific shock common to all countries. The use of panel data in empirical growth regressions has many advantages with respect to cross-sectional 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 solving the inconsistency of empirical estimates which typically arises with omitted country specific effects which, if not uncorrelated with other regressors, lead to a misspecification of the underlying dynamic structure, or with endogenous variables which may be incorrectly treated as exogenous.

There are several issues to be treated in the panel growth regressions literature. Firstly, dependence of the lagged dependent variable and the regressors in $X_{i,t-1}$ with the country-specific fixed effect is allowed in virtually all previous panel studies. In this manner, the country-specific fixed effects are treated as parameters to be estimated and we condition on them, so, their distribution plays no role. This is the so-called fixed effects approach in contrast to the random effects approach that invokes a distribution for η and considers the effects independent of all the regressors in the model. Secondly, Knight et al. (1992) and Islam (1995) among others, have also consider the predetermined nature of the lagged dependent variable with respect to the transitory component of the error term $v_{i,t}$. This point refers to the fact that, by construction, all leads of $y_{i,t-1}$ are correlated with $v_{i,t}$ and, therefore, the within-groups estimator will produce biased estimates in the typical small-T growth panel. In particular, both studies employ the Π-matrix method of Chamberlain (1984). An important drawback of this method is that all the variables in the X vector are considered as strictly exogenous, i.e. all leads and lags of the variables are assumed to be uncorrelated with $v_{i,t}$. This consideration rules out the possibility of feedback from lagged income (i.e. $\ln y$) to current growth determinants such as the rate of investment or the rate of population growth (i.e. the x variables), which seems to be reasonable in the growth context. Finally, Caselli et al. (1996) and Benhabib and Spiegel (2000) among others, take into consideration the predetermined nature[7] of the x variables allowing for the mentioned feedback process. In particular, in order to estimate the model, they use generalized method of moments (GMM) following techniques advanced by Holtz-Eakin et al. (1988) and Arellano and Bond (1991). The assumption that the explanatory variables are predetermined implies a set of moment restrictions that can be used in the context of GMM to generate consistent and efficient estimates of the parameters of interest. More concretely, the employed moment restrictions can be interpreted as an instrumental variables model where lagged levels of the variables are used as instruments for their first-differences. As Blundell and Bond (1998) pointed out, with persistent series such as GDP, lagged levels may be only weak instruments for the equation in first-differences. Thus, in spite of being consistent as N goes to infinity, this estimator is poorly behaved in finite samples. For this reason, these GMM estimates are not very reliable and have not received too much credit in the empirical growth literature. In order to solve this weak-instruments problem, Bond et al. (2001) proposed, in the context of growth regressions, the use of the so-called system-GMM estimator introduced by Arellano and Bover (1995). However, this estimator requires the additional assumption of mean stationarity of the variables. Additional stationarity assumptions for solving this weak-instruments problem are considered an *ad hoc* solution and not very appealing. In the growth regressions framework, this assumption is specially not desiderable since it may be interpreted as assuming that all the countries are in their steady state just after the Second World War.

To the best of my knowledge there is no better alternative to estimate empirical panel growth regressions. The sub-system LIML estimator presented in the previous section is a good candidate for solving the problems described above. First of all, it considers the

[7] This predetermined nature is sometimes denominated weakly exogeneity in the growth literature.

presence of country-specific fixed effects that may be correlated with both lagged income and growth determinants. Secondly, it also takes into consideration the predetermined nature not only of the lagged dependent variable but also of the growth determinants (i.e. feedback from lagged income to current growth determinants is allowed). Thirdly, as it is well-known, LIML estimators alleviate the problem of finite-sample biases caused by weak instruments. Moreover, measurement error considerations can be easily accommodated through additional restrictions on the variance-covariance matrix. On the other hand, it is important to remark that model uncertainty will be considered in the next section.

Given the above, the model to be estimated is given by the following equation and assumption:

$$y_{i,t} = \alpha y_{i,t-1} + \psi x_{i,t-1} + \eta_i + \zeta_t + v_{i,t} \tag{3.10a}$$

$$E\left(v_{i,t} \mid y_i^{t-1}, x_i^{t-1}, \eta_i\right) = 0 \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{3.10b}$$

where $\alpha = 1 + \beta$, $y_{i,t}$ is the GDP per capita for country i in period t, $x_{i,t-1}$ is a $k \times 1$ vector of growth determinants, η_i is a country-specific fixed effect, ζ_t represents a set of time dummies and $v_{i,t}$ is the random disturbance term.

Given current data availability, it is now possible to use 10-year periods in panel growth regressions. This is so because typical sources of "growth data" such as Penn World Tables, cover a broad range of countries over the period 1960 to 2000. By using 10-year periods I aim to avoid the effect of business-cycle fluctuations and, therefore, focus on the long-term growth process. However, I will also present some estimations using 5-year periods data with similar results.

3.4.1 Revisiting the Solow-Swan Model

The baseline empirical growth regression is given by the basic neoclassical growth model, developed by Solow (1956) and Swan (1956). In the empirical counterpart of this model, the vector $x_{i,t-1}$ in (3.10a) includes proxies for the population growth rate (n), the rate of technological progress (g), the rate of depreciation of physical capital (d), and the saving rate (s). In particular, in my regressions, output is measured by GDP per capita at constant 2000 international prices from Penn World Tables 6.2 (PWT62). The saving rate (s) is proxied by the ratio of real domestic investment to GDP from PWT62. Finally, following Mankiw et al. (1992) and Caselli et al. (1996) among others, I choose 0.05 as a reasonable assessment of the value of $g + d$. The Appendix contains more details about the employed data.

I have applied different estimation methods to the Solow-Swan model in two different panel settings, five-year periods and ten-year periods data. The results are presented in Table 3.5. The bulk of the empirical growth regressions literature in based on cross-country OLS regressions as presented in columns (1) and (5). The within-groups (WG) estimator is a slight variant where given the availability of a panel dataset, country dummies can be included in order to allow for the presence of unobserved heterogeneity (i.e. country-specific fixed effects). The results when employing both OLS and WG estimators are in line with previous literature. The problem is that, as previously stated, these estimates are based on the wrong assumptions and thus they are only biased estimates of the real effects. On the other hand, the similarity between WG and diff-GMM estimates is interpreted as an indication of the presence of a weak-instruments problem. This has been previously documented in Bond et al. (2001). As a result, in spite of being based on reasonable assumptions, the diff-GMM estimates are not reliable because they suffer from finite-sample biases.

The sub-system LIML estimation procedure presented in this chapter is applied to the basic Solow-Swan model and the results are shown in columns (4) and (8) of Table 3.5. Inspection of these columns makes it clear the importance of the finite-sample biases in previous differenced GMM estimates of this model. In contrast to previous panel estimates of the rate of convergence using the Solow-Swan framework, I obtain here that the speed of convergence is either low or zero across the countries in the sample. This is true when considering both five-year and ten-year periods. In particular, the point estimate for the convergence rate[8] is roughly zero in both cases. However, the 95% confidence intervals are consistent with convergence rates that vary from −1.5% to 1.1% in the case of five-year periods data and from −2.0% to 1.0% in the case of ten-year data. This result suggests that previous panel studies such as Caselli et al. (1996), where the estimated rate of convergence was surprisingly high, were driven by finite-sample biases. This conclusion will be reinforced in the remaining of the chapter when Barro regressions and model uncertainty will be also taken into account.

By the same token, some differences also arise with respect to other parameter estimates. More concretely, the estimate for $\ln(n_{i,t-1}+g+d)$ is similar in both diff-GMM and sub-system LIML in the sense that they are not significantly different from zero. However, the point estimate is negative in the case of sub-system LIML and positive when using diff-GMM. On the other hand, the estimate of the savings rate coefficient is positive, larger and significant in the case of sub-system LIML but insignificant when using diff-GMM. Moreover, its effect is always larger in the case of ten-year periods data.

3.4.2 Barro Regressions

Since Barro (1991), most of empirical growth regressions are based on a wide variety of specifications given by different variables included in the vector $x_{i,t-1}$ in (3.10a). In this subsection I will apply the sub-system LIML estimator together with OLS, WG and diff-GMM to two distinct panel cross-country growth regressions *a la* Barro. In particular, I focus on the baseline specification of Barro and Lee (1994a) as well as an alternative specification explained below.

The basic empirical framework of Barro regressions with panel data is given by equation (3.10a). Two kind of variables are included in theses regressions, first, initial levels of state variables measured at the beginning of the period (I will now focus on ten-year periods); and second, control or environmental variables, some of which are chosen by governments or private agents. For the baseline specification, as in Barro and Lee (1994a), among the

[8] The convergence rate λ is obtained as follows: $\lambda = \frac{\ln \alpha}{-\tau}$ where τ is either 5 or 10. On the other hand, its standard error is calculated by the delta method.

state variables I include the initial level of per capita GDP, the average number of years of secondary education, and the logarithm of life expectancy. The first is used to proxy the initial stock of physical capital, while the others are proxies for the initial level of human capital in the forms of educational attainment and health. Among the control variables, I include the domestic investment ratio (I/GDP) and the ratio of government consumption to GDP (G/GDP) as in Barro and Lee (1994a). Given data availability in my sample period, the other two control variables are slightly different from those employed in the original specification but they capture similar effects. I consider the price of investment as a measure market prices distortions that exists in the economy and a polity composite index as a proxy of political freedom and stability. GDP, investment share, government consumption, and investment price are taken from PWT62. Secondary education is from Barro and Lee (2000), life expectancy from World Development Indicators 2005 and the polity index from the Polity IV project[9]. In the next section I will explain more about these and other state and control variables.

Table 3.6 shows the results. Columns (1)-(4) refer to the baseline specificacion previously described. In line with Solow-Swan estimation results, the main conclusion from these columns is that the rate of convergence is either very low or zero according to the sub-system LIML estimates. The 95% sub-system LIML confidence interval goes from −1.1% to 1.6%. On the other hand, the conclusions with respect to other explanatory variables may change a lot depending on the estimation method. For instance, investment price has a negative and significative effect on growth according to the sub-system LIML estimates but not according to diff-GMM that suffer from finite-sample bias.

In columns (5)-(8) I present the results from an alternative specification. Imagine a researcher who is testing the effect of democracy on growth. For this purpose, she estimates a growth regression using as state variables the initial level of per capita GDP, the average years of secondary education and the country's population (in millions of people), and as a control variable she decides to only include the domestic investment ratio (I/GDP). There is no clear theoretical justification behind this specification, but neither there is behind the specification in many papers such as Barro and Lee (1994a). Given this specification, the sub-system LIML 95% confidence interval for the convergence rate estimate goes from −0.8% to 2.9%. On the other hand, there are now some results that are different depending not only on the estimation method but also on the specification. For example, in the baseline specification, the effect of the polity index is estimated to be negative and significant while in the alternative specification it is 34% smaller in magnitude and not significant according to the sub-system LIML estimates.

Given the above, it is easy to imagine thousands of Barro regressions in which the convergence parameter estimate will be different across specifications and in which the effects of the explanatory variables will also be different. This would lead us to misleading conclusions even if we consider unbiased and consistent estimates for a given model because we do not know whether this is the correct empirical model or not. This fact illustrates the need to take

[9]A more detailed description of the data sources and variables is in the Appendix

into consideration model uncertainty in empirical growth regressions. In the next section, I combine the sub-system LIML estimates for a given specification with model averaging techniques in order to address model uncertainty.

3.5 Model Uncertainty

I now turn to the issue of model uncertainty which arises because of the lack of clear theoretical guidance on the choice of growth regressors results in a wide set of possible specifications. Therefore, researcher's uncertainty about the value of the parameter of interest in a growth regression exists at distinct two levels. The first one is the uncertainty associated with the parameter conditional on a given empirical growth model. This level of uncertainty is of course assessed in virtually every empirical study. What is not fully assessed is the uncertainty associated with the specification of the empirical growth model. It is typical for a given paper that the specification of the growth regression is taken as essentially known; while some variations of a baseline model are often reported, via different choices of control variables, standard empirical practice does not systematically account for the sensivity of claims about the parameter of interest to model choice.

Many researchers consider that the most promising approach to account for model uncertainty is to employ model averaging techniques to construct parameter estimates that formally address the dependence of model-specific estimates on a given model. In the growth context, Sala-i-Martin et al. (2004) employ the so-called Bayesian Averaging of Classical Estimates (BACE) to determine which growth regressors should be included in linear cross-country growth regressions.[10] In a pure Bayesian spirit, Fernández et al. (2001b) apply the Bayesian Model Averaging approach with different priors but the same objective as Sala-i-Martin et al. (2004). Given that both papers are cross-sectional studies, Moral-Benito (2009a) extends the BACE approach to a panel data setting taking into account the presence of country-specific fixed effects and the endogeneity of the lagged dependent variable. However, there is no paper considering at the same time model uncertainty and the predetermined nature of growth determinants.

Specifically, in this section model averaging techniques are combined with the likelihood-based estimator previously introduced in order to simultaneously address the issues of endogeneity, omitted variable bias, parameter heterogeneity, measurement error and model uncertainty. Thus, we will be able to obtain consistent estimates of what we can call causal effects in the growth context, which take into consideration the dependence of model-specific estimates on a given empirical growth model and, therefore, the uncertainty at the two different levels mentioned above.

[10] See also Raftery (1995).

3.5.1 Growth Determinants

As previously mentioned, the augmented Solow-Swan model can be taken as the baseline empirical growth model. It consists of four determinants of economic growth, initial income, rates of physical and human capital accumulation, and population growth. In addition to those four determinants, Durlauf et al. (2005)'s 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. The set of growth determinants considered in this chapter is only a subset of that identified by Durlauf et al. (2005). This is so because of three main reasons: (i) Data availability in the panel data context for the postwar period 1960-2000 is smaller than in the cross-sectional case. (ii) Since number of models to be estimated increases exponentially with the number of regressors considered and it is necessary to resort to numerical optimization methods for each model estimation, the problem would be computationally intractable if we include too many candidates. (iii) Finally, as found by Ciccone and Jarocinski (2009), the fewer the potential growth determinants considered, the smaller the sensivity of the results. Therefore, for the purpose of robustness, I focus on the subset of available growth determinants given by those variables that are more relevant from a policy maker perspective. This excludes from the analysis geographic variables such as the fraction of land area in geographical tropics, that in spite of being available, they are of little relevance from a policy perspective.

In particular, I consider here the following growth determinants[11]:

- Initial GDP: One of the main features of the neoclassical growth model is the prediction of a low (less than one) coefficient on initial GDP (i.e. it predicts conditional convergence). If the other explanatory variables are held constant, then the economy tends to approach (or not) its long-run position at the rate indicated by the magnitude of the coefficient.

- Investment Ratio: The ratio of investment to output represents the saving rate in the neoclassical growth model. In this model, a higher saving rate raises the steady-state level of output per effective worker and therefore increases the growth rate for a given starting value of GDP. Many empirical studies such as DeLong and Summers (1991) have found an important positive effect of the investment ratio on economic growth.

- Education: In the neoclassical growth model, since the seminal work of Lucas (1988), the concept of capital is usually broadened from physical capital to include human capital. Education is the form of human capital that has generated most of the empirical work. In spite of the positive theoretical effect, many empirical studies have failed in finding such an effect. In particular I consider here the years of secondary education from Barro and Lee (2000).

- Life Expectancy: Another commonly considered form of human capital is health. In particular, the log of life expectancy at birth at the start of each period is typically

[11] A more detailed description of the data and its sources can be found in the Appendix

used as an indicator of health status. There is a growing consensus that improving health can have a large positive impact on economic growth. For example, Gallup and Sachs (2001) argue that wiping out malaria in sub-Saharan Africa could increase per capita GDP growth by 2.6% a year.

- Population Growth: The steady-state level of output per effective worker in the neoclassical growth model is negatively affected by a higher rate of population growth because a portion of the investment is devoted to new workers rather than to raise capital per worker. However, this implication is not always confirmed when estimating empirical growth models.

- Investment Price: Since the seminal work of Agarwala (1983), it is often argued that distortions of market prices impact negatively on economic growth. Given the connection between investment and growth, such market interferences would be especially important if they apply to capital goods. Therefore, following Barro (1991) and Easterly (1993) among others, I consider the investment price level as a proxy for the level of distortions of market prices that exists in the economy.

- Trade Openness: The trade regime/external environment is captured by the degree of openness measured by the trade openness, imports plus exports as a share of GDP. It is often argued that a higher degree of trade openness increases the opportunity set of profitable investments and therefore promotes economic growth. Many authors such as Levine and Renelt (1992) and Frankel and Romer (1999) have considered this ratio.

- Government Consumption: Since the seminal work of Barro (1991), many authors have considered the ratio of government consumption to GDP as a measure of distortions in the economy. The argument is that government consumption has no direct effect on private productivity but lower saving and growth through the distorting effects from taxation or government-expenditure programs.

- Polity Measure: The role of democracy in the process of economic growth has been the source of considerable research effort. However, there is no consensus about how the level of democracy in a country affects economic growth. Some researchers believe that an expansion of political rights (i.e. more democracy) fosters economic rights and tends thereby to stimulate growth. Others think that the growth-retarding aspects of democracy such as the heightened concern with social programs and income redistribution may be the dominant effect. Many authors such as Barro (1996) and Tavares and Wacziarg (2001) have empirically investigated this issue. In this chapter I consider the Polity IV index of democracy/autocracy for analyzing the overall effect of democracy on growth.

- Population: Romer (1987, 1990) and Aghion and Howitt (1992) among others, developed theories of endogenous growth that imply some benefits from larger scale. In particular, if there are significant setup costs at the country level for inventing or adapting new products or production techniques, then the larger economies would, on this ground, perform better. This countrywide scale effect is tested by considering country's population in millions of people.

3.5.2 Bayesian Averaging of Maximum Likelihood Estimates

The basic idea behind model averaging is to estimate the distribution of unknown parameters of interest across different models. The fundamental principle of Bayesian Model Averaging (BMA) 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 and makes inference more reliable.

Formally, consider a generic representation of an empirical model of the form:

$$\Psi = \theta X + \epsilon \tag{3.11}$$

where Ψ is the dependent variable of interest, and X represents a set of covariates. Imagine that there exist potentially very many empirical models, each given by a different combination of explanatory variables (i.e. different vectors X), and each with some probability of being the 'true' model. 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 y -the data-.

In order to obtain parameter estimates that formally consider the dependence of model-specific estimates on a given model, BMA techniques construct point estimates from the posterior distribution of the parameters. This posterior distribution is calculated as a weighted average of all the 2^K model specific posterior distributions. The weights are given by the posterior probability of the model to be the 'true' model[12]. To be more precise, the point estimate of interest will be the mean of the posterior distribution of the parameters given the data:

$$E(\theta|y) = \sum_{j=1}^{2^K} P\left(M_j|y\right) E(\theta|y, M_j)$$

Moreover, if we assume diffuse priors on the parameter space for any given sample size,

[12] A more detailed discussion of the BMA methodology can be found in Hoeting et al. (1999b) and Koop (2003) among others.

or, if we have a large sample for any given prior on the parameter space we can write:[13]

$$E(\theta|y) = \sum_{j=1}^{2^K} P(M_j|y)\, E(\theta|y, M_j) = \sum_{j=1}^{2^K} P(M_j|y)\, \widehat{\theta}^j_{ML} \tag{3.12}$$

where $\widehat{\theta}^j_{ML}$ is the ML estimate for model j. In this particular case, the sub-system LIML estimator presented in Section 4.2. It is important to note at this point, that each of the models being considered here is comprised by a set of simultaneous equations. Therefore, the sub-system LIML estimator maximizes the joint density of all the $1 + 2^K$ variables for all the possible models conditional on the strictly exogenous variables (i.e. initial observations). Then, a regressor is excluded from a particular model by restricting to zero its coefficients in the structural form equation. By doing so, the densities of the different models are comparable.

Similarly, following Leamer (1978) I also compute the posterior variance:

$$\begin{aligned} V(\theta|y) \;&=\; \sum\nolimits_{j=1}^{2^K} P(M_j|y)\, V(\theta|y, M_j) \\ &+\; \sum\nolimits_{j=1}^{2^K} P(M_j|y)\, \left(E(\theta|y, M_j) - E(\theta|y)\right)^2 \end{aligned} \tag{3.13}$$

Inspection of (3.13) shows that the variance incorporates both the estimated variances of the individual models as well as the variance in estimates of the θ's across different models. Hence, the uncertainty at the two different levels mentioned above is taken into account. It is important to note that the posterior mean and the posterior variance considered here are both conditional on the inclusion of a particular regressor in the model. That is to say, when computing both of them from the posterior distribution I will only consider the models in which the coefficient of the regressor is not restricted to be zero (i.e. the model does not include that variable). However, the unconditional posterior mean can be easily obtained by multiplying the conditional posterior mean (column (1) in Table 3.7) times the Posterior Inclusion Probability (PIP) in column 5 of Table 3.7. Similarly, the unconditional posterior variance can be computed according to $V(\theta|y)_{uncond} = [V(\theta|y)_{cond} + E^2(\theta|y)_{cond}] \times PIP - E^2(\theta|y)_{uncond}$.

Moreover, the weights[14] (i.e. the posterior model probabilities $P(M_j|y)$) are based on the Schwarz asymptotic approximation to the Bayes Factor, and therefore:

$$P(M_j|y) = \frac{P(M_j)\,(NT)^{\frac{-k_j}{2}} f(y|\widehat{\theta}_j, M_j)}{\sum_{i=1}^{2^K} P(M_i)\,(NT)^{\frac{-k_i}{2}} f(y|\widehat{\theta}_i, M_i)} \tag{3.14}$$

[13]The equivalence of classical inference and Bayesian inference under diffuse priors is well-known in the classical normal regression model. For the LIML case, Kleibergen and Zivot (2003) show this equivalence for a particular choice of non-informative priors. Note also that the large sample equivalence is only an approximation.

[14]Unweighted counterparts of the three measures in equations (3.12)-(3.13) are not reported here but they are available upon request.

where $f(y|\widehat{\theta}_j, M_j)$ is the maximized likelihood function for model j. Kass and Wasserman (1995) show that the Schwarz asymptotic approximation formula in (3.14) could also be obtained with a reasonable prior on the parameter space[15] that is known as Unit Information Prior (UIP). Moreover, Eicher et al. (2009b) conclude that this UIP combined with the uniform model prior (i.e. all models are equally probable *a priori*) I consider in the chapter outperforms any other possible combination of priors previously considered in the BMA literature in terms of cross-validated predictive performance. This combination of priors also identifies the largest set of growth determinants.

Following Sala-i-Martin et al. (2004) from the posterior distribution I will also estimate the posterior probability that conditional on a variable's inclusion a coefficient has the same sign as its posterior mean (sign certainty probability). The fraction of models that include a particular regressor in which the corresponding t statistic is larger than 2 in absolute value is also reported. Note that this number is not informative about the sign of the estimated coefficient that can be either positive or negative regardless of its significance. Finally, the posterior inclusion probability of a variable is the sum of the posterior probabilities of all models including the variable and it is also reported in Table 3.7. This probability is an indicator of the weighted average goodness-of-fit of models containing a particular variable relative to models not containing that variable. Table 3.7 presents the results when aplying the BAMLE methodology together with the sub-system LIML estimator. Therefore, both model uncertainty and endogeneity are taken into consideration.

Regarding the issue of convergence, the point estimate of the rate of convergence of an economy to its steady state is 0.73%. This estimate is a weighted average of estimates across all possible empirical growth models. However, considering both levels of uncertainty described above (i.e. applying the delta method to the standard error in column (2)), the estimate of the rate of convergence is not significantly different from zero. Therefore I cannot reject the null hypothesis of no conditional convergence across the countries in my sample[16]. This result casts doubt on the conventional wisdom of conditional convergence as a strong empirical regularity in the country level data. For example, early versions of endogenous growth theories (e.g. Romer (1987, 1990) and Aghion and Howitt (1992)) were criticized because in contrast to the neoclassical growth model, they no longer predicted conditional convergence.

The empirical evidence on growth determinants seems to be conclusive for only one variable, the investment ratio. While the associated standard errors are not distributed according to the usual t-distribution, Sala-i-Martin et al. (2004) note that in most cases, having a ratio of posterior mean to standard deviation around two in absolute value indicates an approximate 95-percent Bayesian coverage region that excludes zero. This 'pseudo-t' statistic would indicate that in the case of the investment ratio, its positive effect on growth is significantly different from zero. On the other hand, the probability of its coefficient to be positive is

[15] A prior on the parameter space that is a multivariate normal with mean the MLE of the parameters and variance the inverse of the expected Fisher information matrix for one observation.

[16] This result was previously found in Moral-Benito (2009a), where model uncertainty and the endogeneity of the lagged dependent variable were considered.

100% according to the sign certainty probability. Moreover, in the 98.8% of the estimated models its coefficient was estimated to be significant at the 95% level.

For the rest of the growth determinants the picture emerging from Table 3.7 is a bit pessimistic since little can be said about them once all the potential biases and inconsistencies have been addressed. Based on the mentioned 'pseudo-t' statistic, there is no variable with an estimated causal effect significantly different from zero. At this point it is important to remark the difference between correlations and causal effects. While previous BMA studies applied to growth regressions obtain correlations, I claim to obtain here estimates of what can be labeled as causal effects. This would mean that given the available data, despite of the existence of variables robustly correlated with growth (see for example Sala-i-Martin et al. (2004), Fernández et al. (2001b) and Moral-Benito (2009a)), besides the investment ratio, little can be said about which variables cause economic growth once inference is based on the proper measures of uncertainty.

It is interesting to analyze in more detail these results. There are two possible reasons why the variables do not robustly cause growth according to our results. On the one hand, it might be the case that the coefficients corresponding to a particular variable are very imprecisely estimated in most models despite there is little variation across the estimates in different models. For example, this seems to be the case of the investment price, that has a sign certainty probability of 94.5% but in only 31.3% of the models its coefficient is significantly different from zero. On the other hand, there are other variables whose coefficients are precisely estimated in many models (i.e. high fraction of models with $|tstat| > 2$) but there is a lot of variation across models (low sign certainty probability). This indicates that in some models the coefficient is estimated to be positive and significative and in other models is negative and significative. This is the case of variables such as the polity index for which there is no consensus in the literature about the sign of their effect on growth. In any event, both explanations lead to high posterior variances that preclude the variables from having a robust causal effect on economic growth.

Finally, the posterior inclusion probability (PIP), the sum of the model probabilities of all the models containing a particular variable, is quite high for some variables. For instance, population, which captures scale effects, has a PIP of 98.0%. Therefore, in spite of being not significant, population should be included in empirical growth regressions as a control variable since the models including population are those with the highest probability of being the true empirical growth model (i.e. the models with better goodness-of-fit in relative terms). Other variables with high PIP that should be included are life expectancy and the investment ratio.

For further insights we can see in Figure 3.1 the marginal posterior distributions of the coefficients that correspond to the variables investment share and population. Analogously to the posterior mean, these distributions are weighted averages of marginal posterior distributions conditional on each individual model. More concretely, these posteriors are mixture normal distributions because model-specific posteriors are normal. This is so because we make use of the Bernstein-von Mises theorem[17] (also known as the Bayesian CLT) which basically

[17]Berger (1985) provides an in-depth analysis and an excelent illustration.

states that a Bayesian posterior distribution is well approximated by a normal distribution with mean at the MLE and dispersion matrix equal to the inverse of the Fisher information. BMA marginal posterior distributions consist of two parts, a continuous distribution on the real line and a point mass at zero. Therefore, in addition to the continous mixture normal distribution a gauge that represents the Posterior Inclusion Probability (PIP) of the variables is included in Figure 3.1.

Analyzing Figure 3.1 we can easily observe that despite the investment share's PIP is low, its estimated causal effect on growth is unambiguously positive. This is so because the posterior distribution cumulates more than 99% of its density on the right of zero. On the other hand, zero is clearly outside the classical 95% confidence interval. However, the opposite is true for the population variable. While its PIP is high, its marginal posterior distribution presents probability mass on both sides of zero, indicating that its causal effect on growth could be either positive or negative.

3.6 Concluding Remarks

This chapter has two main contributions: on the one hand, the likelihood-based (or sub-system LIML) counterpart of GMM estimators in a dynamic panel data model with general endogenous or predetermined variables and fixed effects has been introduced and shown to have good (better than its GMM counterpart) finite-sample properties via Monte Carlo simulations. On the other hand, by combining the aforementioned estimator with Bayesian Model Averaging methods, both endogeneity issues and model uncertainty are simultaneously considered in the empirical growth context. To the best of my knowledge, this chapter is the first attempt in doing so.

While both LIML and one-step GMM have approximately the same distribution for sufficiently large sample sizes, based on my Monte Carlo simulations I find that the proposed sub-system LIML estimator outperforms standard GMM in terms of finite-sample behavior. This result can be viewed as a generalization of the single equation case (see for example Anderson et al. (1982)).

Regarding the growth context, my results indicate that both model uncertainty and endogeneity matter in empirical growth regressions. This is so because the conclusions very much depend on whether you consider these issues or not. In particular, I claim that only after addressing both problems we can obtain reliable conclusions about two prominent questions in the empirical growth literature: what variables cause economic growth and, whether there exists conditional convergence or not.

Once model uncertainty and endogeneity issues are controled for, I conclude that the hypothesis of lack of conditional convergence cannot be rejected (at least accross the countries in my sample (see the Appendix)). This result casts doubt on one of the main predictions of the neoclassical model of growth that has been traditionally accepted, the existence of convergence of national economies towards a steady state.

With regard to the causes of economic growth, according to my results, there is only one variable that robustly promotes growth, the investment ratio. This conclusion is based on consistent estimates, and also on the correct measures of uncertainty for inference purposes. As for the rest of growth determinants considered in this chapter, the available empirical evidence is not enough to conclude whether they significantly cause growth or not.

Finally, looking at the posterior inclusion probability of the variables, I conclude that some of them (e.g. population, life expectancy, and the investment ratio) should always be included as controls in empirical growth regressions. This is so because the models that contain these variables are models that have better goodness-of-fit than models without these variables.

3.7 Appendix of Chapter 3

3.7.1 Simultaneous Equations Model (SEM) Representation

In this appendix I present a Simultaneous Equations Model (SEM) representation that allows me to concentrate some free parameters of the resulting log-likelihood in order to make its maximization feasible. The key idea is to translate into the variance-covariance matrix some of the reduced form parameters given the one-to-one mapping between the matrix of coefficients B and the variance-covariance matrix Ω in the FCS representation. As discussed in the main text, this SEM parametrization is a very convenient representation of the model because it allows me to reduce the dimension of the problem by concentrating the log-likelihood of the system with respect to some reduced form parameters.

Given the spirit of the SEM representation, I first define:

$$\eta_i = \gamma_0 y_{i0} + x'_{i1}\gamma_1 + \epsilon_i \tag{3.15}$$

Note that, again, in (3.15) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$ in order to ensure identification of δ.

Moreover, by substituing (3.15) in (3.1) the whole model can be written as follows:

$$y_{i1} = (\alpha + \gamma_0)y_{i0} + x'_{i1}(\beta + \gamma_1) + w'_i\delta + \epsilon_i + v_{i1} \tag{3.16a}$$

and for $t = 2, ..., T$:

$$y_{it} = \alpha y_{i,t-1} + x'_{it}\beta + \gamma_0 y_{i0} + x'_{i1}\gamma_1 + w'_i\delta + \epsilon_i + v_{it} \tag{3.16b}$$

$$x_{it} = \pi_{t0} y_{i0} + \pi_{t1} x_{i1} + \pi^w_t w_i + \xi_{it} \tag{3.16c}$$

where ξ_{it}, γ_1 and π_{t0} are $k \times 1$ vectors, π_{t1} is a $k \times k$ matrix and π^w_t a $k \times m$ matrix.

In order to rewrite the system in matrix form, I define the following $T + (T-1)k \times 1$ vectors of data and errors for individual i:

$$R^S_i = (y_{i1}, y_{i2}, \ldots, y_{iT}, x'_{i2}, x'_{i3}, \ldots, x'_{iT})'$$
$$U_i = (\epsilon_i + v_{i1}, \ldots, \epsilon_i + v_{iT}, \xi'_{i2}, \ldots, \xi'_{iT})'$$

Therefore I am now able to rewrite the model in matrix form as follows:

$$B^S R^S_i = \Pi z_i + U_i \tag{3.17}$$

where B^S and Π are matrices of coefficients defined below and z_i is the $(1+k+m) \times 1$ vector of strictly exogenous variables:

$$z_i = (y_{i0}, x'_{i1}, w'_i)'$$

Moreover, if I additionally define the following vectors:

$$\begin{aligned} R_{i1}^S &= (y_{i1}, y_{i2}, \ldots, y_{iT})' \\ R_{i2}^S &= (x_{i2}', x_{i3}', \ldots, x_{iT}')' \\ U_{i1} &= (\epsilon_i + v_{i1}, \ldots, \epsilon_i + v_{iT})' \\ U_{i2} &= (\xi_{i2}', \ldots, \xi_{iT}')' \end{aligned}$$

it is then possible to rewrite:

$$\left(\begin{array}{c|c} B_{11}^S & B_{12}^S \\ \hline 0 & I_{k-1} \end{array}\right) \left(\begin{array}{c} R_{i1}^S \\ \hline R_{i2}^S \end{array}\right) = \left(\begin{array}{c} \Pi_1 \\ \hline \Pi_2 \end{array}\right) z_i + \left(\begin{array}{c} U_{i1} \\ \hline U_{i2} \end{array}\right) \tag{3.18}$$

where:

$$B_{11}^S = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ -\alpha & 1 & 0 & \ldots & 0 \\ 0 & -\alpha & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -\alpha & 1 \end{pmatrix}_{T\times T} \qquad B_{12}^S = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ -\beta' & 0 & \ldots & 0 \\ 0 & -\beta' & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & -\beta' \end{pmatrix}_{T\times k(T-1)}$$

$$\Pi_1 = \begin{pmatrix} \alpha + \gamma_0 & \beta' + \gamma_1' & \delta' \\ \gamma_0 & \gamma_1' & \delta' \\ \vdots & \vdots & \vdots \\ \gamma_0 & \gamma_1' & \delta' \end{pmatrix}_{T\times(1+k+m)} \qquad \Pi_2 = \begin{pmatrix} \pi_{20} & \pi_{21} & \pi_2^w \\ \vdots & \vdots & \vdots \\ \pi_{T0} & \pi_{T1} & \pi_T^w \end{pmatrix}_{k(T-1)\times(1+k+m)}$$

In contrast to the FCS representation, considering the SEM parametrization we can see that the number of non-zero coefficients in the matrix B^S is only $k+1$. This is so because they have been "translated" into the variance-covariance matrix of the model that is no longer block-diagonal. In particular:

$$\Omega^S = Var(U_i) = Var\left(\begin{array}{c} U_{i1} \\ \hline U_{i2} \end{array}\right) = \left(\begin{array}{c|c} \Omega_{11}^S & \Omega_{12}^S \\ \hline \Omega_{21}^S & \Omega_{22}^S \end{array}\right) \tag{3.19}$$

where:

- Ω_{11}^S has the classical error-component form but allowing for time-series heteroskedasticity:

  $$\Omega_{11}^S = \sigma_\epsilon^2 \iota\iota' + \begin{pmatrix} \sigma_{v_1}^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{v_T}^2 \end{pmatrix}$$

  where ι is a $T \times 1$ vector of ones.

- Ω_{22}^S is the $(T-1)k \times (T-1)k$ covariance matrix that gathers all the contemporaneous and dynamic relationships between the x variables:

$$\Omega_{22}^S = \begin{pmatrix} \Sigma_{2,2} & & & \\ \Sigma_{2,3} & \Sigma_{3,3} & & \\ \vdots & \vdots & \ddots & \\ \Sigma_{2,T} & \Sigma_{3,T} & \dots & \Sigma_{T,T} \end{pmatrix}$$

where $\Sigma_{f,g}$ is the $k \times k$ covariance matrix between x_{if} and x_{ig}.

- Ω_{12}^S captures the feedback process. In particular, given the assumptions above I can write:

$$cov(\epsilon_i, \xi_{it}) = \phi_t \qquad \forall t = 2, ..., T \tag{3.20a}$$

$$cov(v_{ih}, \xi_{it}) = \begin{cases} \psi_{h,t} & \text{if } h < t \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{3.20b}$$

where ϕ_t, $\psi_{h,t}$ and $\mathbf{0}$ are $k \times 1$ vectors. Therefore:

$$\Omega_{12}^S = \begin{pmatrix} \phi'_2 + \psi'_{1,2} & \phi'_3 + \psi'_{1,3} & \dots & \phi'_T + \psi'_{1,T} \\ \phi'_2 & \phi'_3 + \psi'_{2,3} & \dots & \phi'_T + \psi'_{2,T} \\ \phi'_2 & \phi'_3 & \dots & \phi'_T + \psi'_{3,T} \\ \vdots & \vdots & \ddots & \vdots \\ \phi'_2 & \phi'_3 & \dots & \phi'_T + \psi'_{T-1,T} \\ \phi'_2 & \phi'_3 & \dots & \phi'_T \end{pmatrix}_{T \times (T-1)k}$$

Under normal errors the log-likelihood for the model can be written as[18]:

$$L_S \propto -\frac{N}{2} \ln \det(\Omega^S) - \frac{1}{2} tr\left((\Omega^S)^{-1} U'U\right) \tag{3.21}$$

where U' is a $T + (T-1)k \times N$ matrix that consists of the U_i column vectors of each of the N individuals. Note that this is an integrated likelihood that is marginal on η_i but conditional on $z_i = (y_{i0}, x'_{i1}, w_i)'$:

$$f(y_i^T, x_i^T | z_i) = \int \prod_{t=1}^{T} f(y_{it} | y_i^{t-1}, x_i^t, w_i, \eta_i) \prod_{t=2}^{T} f(x_{it} | y_i^{t-1}, x_i^{t-1}, w_i, \eta_i) dG(\eta_i | z_i) \tag{3.22}$$

As in the case of the FCS representation in the main text, the maximizer of L_S is a consistent and asymptotically normal estimator regardless of non-normality. Moreover, the number of parameters to be estimated in (3.21) is the same as in (3.4). In order to make the problem feasible I will work with the concentrated log-likelihood with respect to the free parameters in the matrices Π_2 and Ω_{22}^S (i.e. the parameters that capture the dynamic and

[18]Note that $\det(B^S) = 1$

contemporaneous relationships between the x variables and between the variables and the strictly exogenous variables). See the Appendix for more details on the concentration of the SEM log-likelihood.

3.7.2 Concentrated Likelihood using the SEM Parametrization

Maximizing the log-likelihood in (3.21) may be cumbersome (or even impossible depending on the number of available observations) since the dimension of the numerical optimization problem is enormous. In particular, the number of parameters to be estimated (p) in (3.21) is determined by the following expression:

$$p = 3 + 2k + T + (T-1)(2+k+m)k + \frac{(T-1)k[(T-1)k+1]}{2} + \sum_{r=1}^{T-1} rk$$

As an illustrative example, suppose we have a panel with $T = 5$, $k = 7$ and $m = 4$, then $p = 862$. This number is huge and may cause the problem to be intractable, but it can be drastically reduced by concentrating some free parameters of the model. In particular, for this illustrative example, the number of parameters after concentrating the log-likelihood is reduced from $p = 862$ to $p = 120$.

The log-likelihood function in (3.21) will be concentrated with relation to Ω_{22}^S and Π_2 under the assumption that both terms are unconstrained. The concentrated log-likelihood will then be maximized by means of numerical optimization with relation to B_{11}^S, B_{12}^S, Π_1, Ω_{11}^S and Ω_{12}^S that are all restricted. In what follows, I refer to Ω_{22}^S, B_{11}^S, B_{12}^S, Ω_{11}^S and Ω_{12}^S as Ω_{22}, B_{11}, B_{12}, Ω_{11} and Ω_{12} for the sake of notational simplicity.

By grouping the observations for all individuals in columns, the model can be written as follows:

$$\left(\begin{array}{c:c} B_{11} & B_{12} \\ \hdashline 0 & I_{k-1} \end{array}\right) \left(\begin{array}{c} R_1' \\ \hdashline R_2' \end{array}\right) = \left(\begin{array}{c} \Pi_1 \\ \hdashline \Pi_2 \end{array}\right) Z' + \left(\begin{array}{c} U_1' \\ \hdashline U_2' \end{array}\right)$$

First of all, we define:

$$\begin{aligned} \Omega^{-1} &= \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}^{-1} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \\ F_{12} &= G_{12} G_{22}^{-1} \\ F_{21} &= F_{12}' \end{aligned}$$

and then rewrite:

$$\begin{aligned} \det \Omega &= \det \Omega_{11} / \det G_{22} \\ tr(\Omega^{-1} U' U) &= tr(\Omega_{11}^{-1} U_1' U_1) + 2tr(G_{12} U_2' U_1) + tr(G_{22} U_2' U_2) + tr(G_{12} G_{22}^{-1} G_{21} U_1' U_1) \end{aligned}$$

Therefore, (3.21) can be written as follows:

$$\begin{aligned} L \;\propto\; & -\frac{N}{2} \ln \det \Omega_{11} + \frac{N}{2} \ln \det G_{22} - \frac{1}{2} tr(\Omega_{11}^{-1} U_1' U_1) - tr(F_{12} G_{22} U_2' U_1) \qquad (3.23) \\ & - \frac{1}{2} tr(G_{22} U_2' U_2) - \frac{1}{2} tr(F_{12} G_{22} F_{21} U_1' U_1) \end{aligned}$$

Note that we can also write $\Omega_{11}^{-1} = G_{11} - G_{12}G_{22}^{-1}G_{21}$ and I have added and substracted the term $tr(G_{12}G_{22}^{-1}G_{21}U_1'U_1)$.

Step 1: Concentrating out Π_2

Noting that $U_2' = R_2' - \Pi_2 Z'$, we can maximize the likelihood in (3.23) with respect to Π_2 and obtain its ML estimate:

$$\widehat{\Pi}_2 = R_2'Z(Z'Z)^{-1} + F_{21}U_1'Z(Z'Z)^{-1}$$

Given $\widehat{\Pi}_2$ we can write:

$$\begin{aligned}\widehat{U}_2'U_1 &= R_2'QU_1 - F_{21}U_1'MU_1 \\ \widehat{U}_2'\widehat{U}_2 &= R_2'QR_2 + F_{21}U_1'MU_1F_{12}\end{aligned}$$

where M is the projection matrix on the exogenous variables of the system and Q the annihilator:

$$\begin{aligned}M &= Z(Z'Z)^{-1}Z' \\ Q &= I_N - M\end{aligned}$$

Replacing in (3.23), we obtain L_2, the log-likelihood concentrated with respect to Π_2:

$$\begin{aligned}L_2 &\propto -\frac{N}{2}\ln\det\Omega_{11} + \frac{N}{2}\ln\det G_{22} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) \\ &- \frac{1}{2}tr\{(R_2 + U_1F_{12})'Q(R_2 + U_1F_{12})G_{22}\}\end{aligned} \tag{3.24}$$

Step 2: Concentrating out Ω_{22}

I now turn to the concentration of L_2 with relation to Ω_{22}. Note that the log-likelohood is now written in terms of G_{22} and therefore, in practice I will obtain the concentrated likelihood with respect to G_{22} instead of Ω_{22}. However, since they are unconstrained, this is simply a matter of notation.

First, we define:

$$H = (R_2 + U_1F_{12})'Q(R_2 + U_1F_{12})$$

Therefore:

$$L_2 \propto -\frac{N}{2}\ln\det\Omega_{11} + \frac{N}{2}\ln\det G_{22} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) - \frac{1}{2}tr\{HG_{22}\}$$

By differentiating the log-likelihood function, we obtain:

$$\begin{aligned}\mathbf{d}L_2 &= \frac{N}{2}tr(G_{22}^{-1}\mathbf{d}G_{22}) - \frac{1}{2}tr(H\mathbf{d}G_{22}) \\ &= tr[(\frac{N}{2}G_{22}^{-1} - \frac{1}{2}H)\mathbf{d}G_{22}] = 0\end{aligned}$$

This implies that:

$$\widehat{G}_{22}^{-1} = \frac{1}{N}H$$

and so the final concentrated log-likelihood is:

$$L_3 \propto -\frac{N}{2}\ln\det\Omega_{11} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) - \frac{N}{2}\ln\det(\frac{1}{N}H) \tag{3.25}$$

3.7.3 Monte Carlo Details

For simulating the data in the Monte Carlo experiment, I first estimate a trivariate VAR process for GDP[19] (y), investment ratio (x^1) and population growth (x^2). In particular, I consider the following VAR process:

$$w_{it} = \Gamma w_{it-1} + \zeta_i + \vartheta_{it}$$

where:

$$\begin{aligned} w_{it} &= (y_{it}, x_{it}^1, x_{it}^2)' \\ \zeta_i &= (\zeta_i^y, \zeta_i^1, \zeta_i^2)' \\ \vartheta_{it} &= (\epsilon_{it}^y, \epsilon_{it}^1, \epsilon_{it}^2)' \\ Var\left((w_{i0}', \zeta_i')'\right) &= \Omega_{MC} \\ Var\left(\vartheta_{it}\right) &= \Sigma_{MC} \end{aligned}$$

Once I get the estimates $\widehat{\Gamma}$, $\widehat{\Omega}_{MC}$ and $\widehat{\Sigma}_{MC}$, the procedure for generating the data is as follows:

1. Generate w_{i0} and ζ_i according to $(w_{i0}', \zeta_i')' \sim N(0, \widehat{\Omega}_{MC})$.

2. For $t = 1, ..., T$:

   (a) Generate ϑ_{it} according to $\vartheta_{it} \sim N(0, \widehat{\Sigma}_{MC})$

   (b) Then generate w_{it} according to $w_{it} = \widehat{\Gamma} w_{it-1} + \zeta_i + \vartheta_{it}$

More concretely, the employed parameter values when considering ten-year periods in the baseline Monte Carlo simulations are as follows:

$$\widehat{\Gamma} = \begin{pmatrix} .95 & .20 & -.10 \\ .10 & .70 & 0 \\ -.20 & 0 & .60 \end{pmatrix} \qquad \widehat{\Sigma}_{MC} = \begin{pmatrix} .167 & & \\ -.002 & .071 & \\ -.002 & .002 & .077 \end{pmatrix}$$

$$\widehat{\Omega}_{MC} = \begin{pmatrix} .913 & & & & & \\ .367 & .602 & & & & \\ -.061 & -.039 & .021 & & & \\ -.095 & -.088 & .007 & .019 & & \\ -.010 & .051 & -.002 & -.007 & .017 & \\ .161 & .072 & -.004 & -.018 & .0005 & .034 \end{pmatrix}$$

As mentioned in the main text, additional Monte Carlo experiments were carried out considering five-year periods data for the calibration. In this case I obtain and use the

[19] In the estimation of the VAR all variables are expressed in logs

following parameter values:

$$\widehat{\Gamma} = \begin{pmatrix} .98 & .10 & -.05 \\ .05 & .80 & 0 \\ .10 & 0 & .40 \end{pmatrix} \qquad \widehat{\Sigma}_{MC} = \begin{pmatrix} .125 & & \\ -.001 & .109 & \\ -.001 & .0003 & .085 \end{pmatrix}$$

$$\widehat{\Omega}_{MC} = \begin{pmatrix} .913 & & & & & \\ .400 & .657 & & & & \\ -.049 & -.029 & .019 & & & \\ -.089 & -.119 & .009 & .027 & & \\ -.132 & -.031 & .007 & .005 & .024 & \\ .166 & .086 & -.007 & -.019 & -.023 & .032 \end{pmatrix}$$

Moreover, the Monte Carlo results of the five-year periods experiments are represented in the following table:

Table 3.1: Additional Monte Carlo Results

	$\alpha = 0.98$			$\beta_1 = 0.10$			$\beta_2 = -0.05$		
	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML
				Panel A: $T = 4, N = 100$					
median	.453	.357	.952	.064	−.096	.075	−.059	−.017	−.052
iqr	.078	.342	.169	.082	.231	.194	.096	.160	.173
MAE	.527	.623	.076	.048	.198	.102	.047	.082	.086
				Panel B: $T = 4, N = 500$					
median	.454	.508	.970	.063	−.076	.087	−.061	−.023	−.051
iqr	.034	.317	.122	.036	.147	.099	.042	.075	.073
MAE	.526	.472	.052	.037	.176	.049	.022	.041	.037
				Panel C: $T = 4, N = 1000$					
median	.454	.609	.973	.062	−.039	.091	−.061	−.029	−.054
iqr	.024	.288	.108	.027	.116	.067	.031	.057	.058
MAE	.526	.371	.048	.038	.139	.034	.017	.032	.029
				Panel D: $T = 8, N = 100$					
median	.712	.650	.982	.081	.013	.098	−.055	−.025	−.050
iqr	.043	.145	.089	.046	.101	.096	.065	.080	.088
MAE	.268	.330	.043	.027	.088	.048	.031	.044	.044
				Panel E: $T = 8, N = 500$					
median	.714	.761	.982	.079	.029	.099	−.059	−.028	−.051
iqr	.019	.112	.067	.024	.052	.041	.028	.038	.036
MAE	.266	.219	.034	.021	.071	.021	.016	.025	.018
				Panel F: $T = 8, N = 1000$					
median	.714	.826	.979	.081	.050	.101	−.060	−.035	−.051
iqr	.013	.084	.056	.016	.039	.032	.019	.025	.025
MAE	.266	.154	.027	.019	.050	.016	.012	.018	.013

Notes: 1,000 replications. iqr is the 75th-25th interquartile range; MAE denotes the median absolute error. Parameter values calibrated to five-year periods data.

3.7.4 Data Appendix

Table 3.2: Variable Definitions and Sources

Variable	Source	Definition
GDP	PWT 6.2	Logarithm of GDP per capita (2000 US dollars at PP)
I/GDP	PWT 6.2	Ratio of real domestic investment to GDP
Education	Barro and Lee (2000)	Stock of years of secondary education in the total population
Pop. Growth	PWT 6.2	Average growth rate of population
Population	PWT 6.2	Population in millions of people
Inv. Price	PWT 6.2	Purchasing-power-parity numbers for investment goods
Trade Openness	PWT 6.2	Exports plus imports as a share of GDP
G/GDP	PWT 6.2	Ratio of government consumption to GDP
ln (life expect)	WDI 2005	Logarithm of the life expectancy at birth
Polity	Polity IV Project	Composite index given by the democracy score minus the autocracy score. Original range -10,-9,...,10, normalized 0-1.

Notes: All variables are available for all the countries in the sample (see table below) and for the whole period 1960-2000. PWT 6.2 refers to Penn World Tables 6.2 and it can be fount at http://pwt.econ.upenn.edu/. WDI 2005 refers to World Development Indicators 2005. Data from Barro and Lee (2000) is available at http://www.cid.harvard.edu/ciddata/ciddata.html. Finally, data from the Polity IV Project can be downloaded from http://www.systemicpeace.org/polity/polity4.htm.

Table 3.3: List of Countries in Chapter 3

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

3.7.5 Figures

Figure 3.1: Posterior Distributions of Selected Coefficients

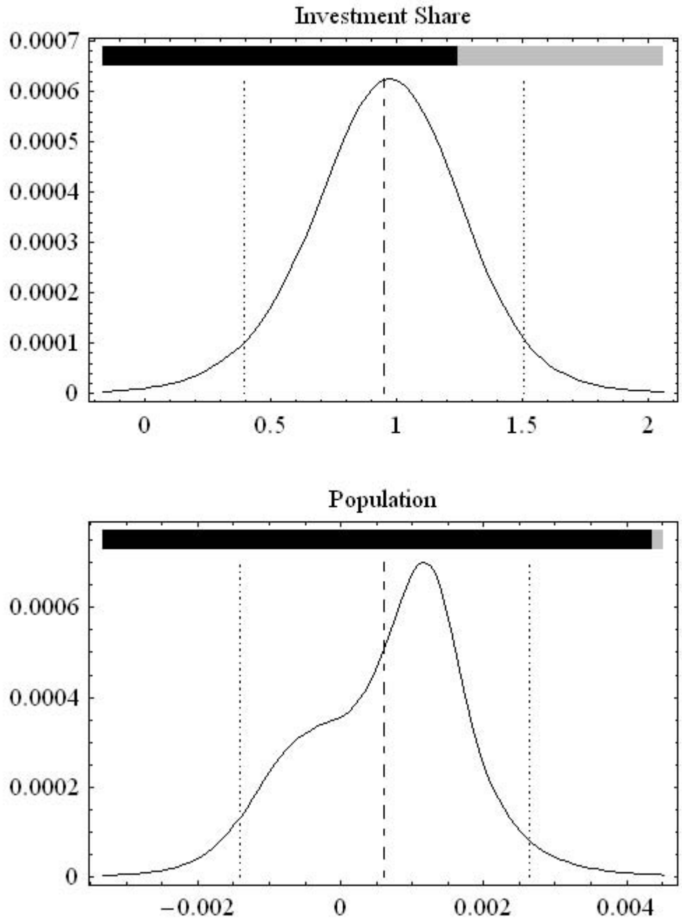


Figure 3.1 presents the marginal posterior distributions of the investment share and population coefficients. In particular, each graph consists of two parts: a gauge on top of the graphs that indicates the Posterior Inclusion Probability (PIP) of the variables and the normal mixture density for each coefficient. A dashed vertical line indicates the posterior mean conditional on inclusion presented in column 1 of Table 3.7. The equivalent to a classical 95% confidence interval is represented by two vertical dotted lines.

3.7.6 Tables

Table 3.4: Monte Carlo Results

	$\alpha = 0.95$			$\beta_1 = 0.20$			$\beta_2 = -0.10$		
	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML
				Panel A: $T = 4, N = 100$					
median	.426	.440	.900	.084	$-.118$	.154	$-.097$	$-.155$	$-.107$
% bias	55.2%	53.7%	5.2%	57.8%	159.0%	23.2%	2.6%	55.0%	6.6%
iqr	.079	.319	.157	.100	.265	.205	.096	.172	.161
MAE	.524	.510	.070	.116	.320	.113	.047	.091	.081
				Panel B: $T = 4, N = 500$					
median	.432	.691	.929	.083	.022	.173	$-.096$	$-.133$	$-.102$
% bias	54.5%	27.2%	2.2%	58.4%	88.8%	13.4%	4.4%	32.9%	2.3%
iqr	.033	.238	.104	.046	.172	.108	.046	.071	.070
MAE	.518	.260	.038	.117	.181	.056	.023	.042	.035
				Panel C: $T = 4, N = 1000$					
median	.432	.789	.932	.084	.089	.179	$-.096$	$-.120$	$-.103$
% bias	54.6%	16.9%	1.9%	57.9%	55.5%	10.6%	4.0%	20.4%	3.4%
iqr	.025	.176	.092	.035	.135	.080	.034	.052	.049
MAE	.518	.164	.032	.116	.116	.042	.017	.028	.024
				Panel D: $T = 8, N = 100$					
median	.685	.730	.935	.154	.074	.184	$-.112$	$-.151$	$-.102$
% bias	27.8%	23.1%	1.5%	23.2%	63.0%	7.8%	11.5%	51.0%	2.3%
iqr	.044	.111	.073	.062	.114	.124	.069	.086	.090
MAE	.265	.220	.035	.049	.126	.061	.035	.058	.045
				Panel E: $T = 8, N = 500$					
median	.687	.867	.947	.150	.143	.194	$-.114$	$-.124$	$-.102$
% bias	27.7%	8.7%	.4%	25.2%	28.6%	3.0%	14.5%	23.9%	2.3%
iqr	.021	.057	.046	.031	.057	.054	.028	.040	.039
MAE	.263	.083	.021	.050	.057	.027	.018	.028	.019
				Panel F: $T = 8, N = 1000$					
median	.687	.903	.949	.152	.169	.197	$-.116$	$-.115$	$-.102$
% bias	27.7%	4.9%	.1%	23.8%	15.7%	1.4%	16.0%	14.6%	2.3%
iqr	.014	.043	.036	.021	.044	.041	.020	.028	.026
MAE	.263	.047	.017	.048	.033	.021	.016	.018	.013

Notes: 1,000 replications. % bias gives the percentage median bias for all the estimates; iqr is the 75th-25th interquartile range; MAE denotes the median absolute error. Parameter values calibrated to ten-year periods data.

Table 3.5: Solow-Swan Model Estimation Results

	Five-year data				Ten-year data			
	OLS	WG	diff GMM	sub-sys LIML	OLS	WG	diff GMM	sub-sys LIML
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is $\ln(y_{i,t})$							
$\ln(y_{i,t-1})$	0.963	0.843	0.830	1.012	0.927	0.718	0.717	1.025
	(0.007)	(0.025)	(0.050)	(0.034)	(0.014)	(0.050)	(0.112)	(0.076)
$\ln(s_{i,t-1})$	0.088	0.091	0.035	0.095	0.167	0.166	0.009	0.222
	(0.010)	(0.018)	(0.034)	(0.022)	(0.019)	(0.036)	(0.085)	(0.049)
$\ln(n_{i,t-1}+g+d)$	−0.204	−0.137	0.128	0.020	−0.441	−0.327	0.557	−0.102
	(0.041)	(0.071)	(0.108)	(0.082)	(0.085)	(0.163)	(0.325)	(0.342)
Implied λ	0.007	0.034	0.037	−0.002	0.008	0.033	0.033	−0.003
	(0.001)	(0.006)	(0.012)	(0.007)	(0.002)	(0.007)	(0.016)	(0.009)
Observations	584	584	511	584	292	292	219	292
Countries	73	73	73	73	73	73	73	73

Notes: In all columns a set of time dummies is included in the regressions. Columns (1) and (5) refer to the OLS estimation without country-specific fixed effects and all regressors considered as exogenous. In columns (2) and (6) the within-group estimator is employed and therefore fixed effects are included. However all regressors are assumed to be strictly exogenous. Finally, columns (3)-(4) and (7)-(8) present different estimates of the Solow-Swan version of the model in (3.10a)-(3.10b), where both fixed effects and weakly exogeneity are considered. In particular, columns (3) and (7) refer to the differenced GMM estimation and columns (4) and (8) present the estimation results when using the sub-system LIML estimator presented in Section 4.2. Standard errors are in parenthesis.

Table 3.6: Barro Regressions Estimation Results

	Baseline Specification Ten-year data				Alternative Specification Ten-year data			
	OLS	WG	diff GMM	sub-sys LIML	OLS	WG	diff GMM	sub-sys LIML
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is $\ln(y_t)$							
$\ln(y_{t-1})$	0.845	0.683	0.842	0.977	0.971	0.624	0.438	0.899
	(0.021)	(0.052)	(0.075)	(0.068)	(0.019)	(0.051)	(0.107)	(0.084)
Education	0.040	0.039	0.055	0.030	0.016	0.036	0.076	0.030
	(0.015)	(0.036)	(0.081)	(0.056)	(0.017)	(0.032)	(0.046)	(0.054)
ln (life expect)	0.829	0.478	0.709	0.862				
	(0.108)	(0.224)	(0.488)	(0.190)				
I/GDP	0.588	0.781	0.857	1.114	0.891	0.797	0.351	1.268
	(0.133)	(0.213)	(0.279)	(0.244)	(0.132)	(0.193)	(0.284)	(0.293)
G/GDP	−0.246	−0.465	−0.314	−0.546				
	(0.115)	(0.284)	(0.534)	(0.318)				
Inv. Price	−0.0004	−0.0007	−0.0008	−0.0010				
	(0.0002)	(0.0003)	(0.0006)	(0.0004)				
Polity	−0.042	−0.201	−0.260	−0.256	0.054	−0.167	−0.338	−0.169
	(0.041)	(0.061)	(0.083)	(0.084)	(0.042)	(0.058)	(0.082)	(0.096)
Population					0.0003	0.017	0.020	0.0012
					(0.0001)	(0.0003)	(0.0003)	(0.0004)
Implied λ	0.017	0.038	0.017	0.002	0.003	0.047	0.082	0.011
	(0.003)	(0.008)	(0.009)	(0.007)	(0.002)	(0.008)	(0.024)	(0.009)
Observations	292	292	219	292	292	292	219	292
Countries	73	73	73	73	73	73	73	73

Notes: The baseline specification is the same as in Barro and Lee (1994a) and the alternative specification is explained in the main text. In all columns a set of time dummies is included in the regressions. Columns (1) and (5) refer to the OLS estimation without country-specific fixed effects and all regressors considered as exogenous. In columns (2) and (6) the within-group estimator is employed and therefore fixed effects are included. However all regressors are assumed to be strictly exogenous. Finally, columns (3)-(4) and (7)-(8) present different estimates of two versions of the model in (3.10a)-(3.10b) where both fixed effects and weakly exogeneity are considered. In particular, columns (3) and (7) refer to the differenced GMM estimation and columns (4) and (8) present the estimation results when using the sub-system LIML estimator presented in Section 4.2. Standard errors are in parenthesis.

Table 3.7: BAMLE Results

	Posterior mean conditional on inclusion (1)	Posterior s.d. conditional on inclusion (2)	Sign certainty probability (3)	Fraction of models with $\lvert tstat \rvert > 2$ (4)	Posterior Inclusion Probability (5)
	Dependent variable is $\ln(y_t)$				
$\ln(y_{t-1})$	0.930	0.091	100.0%	100.0%	-
I/GDP	0.949	0.284	100.0%	98.8%	63.4%
Education	0.033	0.058	75.4%	4.3%	56.1%
Pop. Growth	−0.566	2.897	57.8%	17.6%	55.3%
Population	0.0006	0.0010	79.3%	14.1%	98.0%
Inv. Price	−0.0005	0.0006	94.5%	31.3%	47.9%
Trade Openness	0.038	0.052	87.1%	64.1%	60.7%
G/GDP	0.048	0.204	60.9%	25.0%	60.3%
ln(life expect)	0.078	0.222	78.1%	60.9%	75.7%
Polity	−0.125	0.128	68.4%	46.9%	50.4%

Notes: In this table, the sub-system LIML estimator introduced in Section 4.2 is combined with the BAMLE methodology described in the main text. The sample covers the period 1960 to 2000 divided in 10-years subperiods. Column (1) reports the weighted average of the sub-system LIML estimates across all the possible models containing the variable (i.e. it corresponds to equation (3.12)). Column (2) refers to the square root of the posterior variance presented in equation (3.13). In column (3) I report the sign certainty probability. Column (4) presents the percentage of models in which the coefficient is significantly different from zero (positive or negative). Finally, column (5) refers to the posterior inclusion probability of a variable to be included in the 'true' empirical growth model. It is calculated as the sum of the posterior model probabilities of all the models containing that variable. Finally, while the results on the table are based on the assumption of a prior expected model size equal to $K/2$ (i.e. uniform model prior), results with different prior expected model sizes are very similar and available upon request.

Concluding Remarks

Model uncertainty and endogeneity issues hamper consensus on the key determinants of economic growth and the convergence debate. The main objective of this dissertation is to shed light on these two debates by simultaneously addressing endogeneity and model uncertainty in the context of growth regressions.

Model uncertainty emerges because theory does not provide enough guidance to select the proper empirical model. Many researchers consider that the most promising approach to accounting for model uncertainty is to employ model averaging techniques to construct parameter estimates that formally address the dependence of model-specific estimates on a given model. This dissertation contributes to the literature on model averaging by extending this methodology to panel data settings.

Endogeneity arises because growth regressors are typically correlated with unobservables influencing economic growth. The use of panel data is one natural alternative to overcome the endogeneity issue. This dissertation contributes to the literature on panel data and endogeneity by proposing a novel panel maximum likelihood estimator with endogenous regressors that is equivalent to first differenced GMM in terms of assumptions but strongly preferred in terms of finite sample performance.

Chapter 1 presents a survey of model averaging methods with special emphasis on economic applications. Chapter 1 also highlights an important line of open research which is pursued in this dissertation, the extension of model averaging techniques to settings with endogenous regressors and panel data. An interesting line for further research that is in my research agenda is the combination of model averaging with treatment effects estimators such as matching estimators in which model uncertainty is also present in the choice of conditioning variables.

Chapter 2 extends the model averaging apparatus to panel data models with an application to growth empirics. There are in the literature several papers applying model averaging to cross-country growth regressions with cross-sectional data, but the extension to dynamic panels with country-specific effects is not straightforward. I propose in Chapter 2 several alternatives. The empirical findings of Chapter 2 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. However, given the exogeneity assumption for all the growth regressors considered, the estimates in this chapter should not be interpreted as causal effects.

Chapter 3 proposes a likelihood function for dynamic panel data models with predetermined or endogenous variables and fixed effects. The resulting maximum likelihood estimator can be interpreted as the LIML counterpart of panel GMM estimators. Via Monte Carlo simulations, I conclude in this chapter that the finite-sample performance of the proposed estimator is better than that of the commonly-used standard GMM. Therefore, the proposed estimator is a relevant contribution per se. Since regressors are assumed to be endogenous in Chapter 3, the estimates resulting from the LIML estimator introduced in this chapter can be interpreted as causal effects. Moreover, given the availability of this likelihood function, I also combine the estimator with model averaging techniques in order to simultaneously overcome endogeneity and model uncertainty.

In contrast to the previous consensus in the empirical growth literature, empirical results indicate that once endogeneity and model uncertainty are accounted for at the same time, the estimated convergence rate is not significantly different from zero. Moreover, there seems to be only one variable, the investment ratio, that causes long-run economic growth.

Bibliography

Agarwala, R., "Price Distortions and Growth in Developing Countries," *World Bank staff working papers*, 1983, *No. 575.*

Aghion, P. and P. Howitt, "A Model of Endogenous Growth through Creative Destruction," *Econometrica*, 1992, *60*, 323–351.

Ahn, S. and P. Schmidt, "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 1995, *68*, 5–27.

Alonso-Borrego, C. and M. Arellano, "Symmetrically Normalized Instrumental - Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 1999, *17*, 36–49.

Alvarez, J. and M. Arellano, "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 2003, *71*, 1121–1159.

Anderson, T., N. Kunitomo, and T. Sawa, "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 1982, *50*, 1009–1027.

Angrist, J. and J. Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics," *The Journal of Economic Perspectives*, 2010, *24*, 3–30.

Arellano, M., *Panel Data Econometrics*, Oxford, U.K.: Oxford University Press, 2003.

_ **and O. Bover**, "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, 1995, *68*, 29–52.

_ **and S.R. Bond**, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 1991, *58*, 277–297.

Avramov, D., "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 2002, *64*, 423–258.

Banerjee, M. and E. Frees, "Influence Diagnostics for Linear Longitudinal Models," *Journal of the American Statistical Association*, 1997, *92*, 999–1005.

Barnard, G., "New Methods of Quality Control," *Journal of the Royal Statistical Society. Series A (General)*, 1963, *126*, 255–258.

Barro, R., "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, 1991, *106* (2), 407–443.

_ , "Democracy and Growth," *Journal of Economic Growth*, 1996, *1*, 1–27.

_ **and J. Lee**, "Losers and Winners in Economic Growth," *Proceedings of the World Bank Annual Conference on Development Economics*, 1994, *Washington D.C.*, 267–297.

_ **and** _ , "Sources of Economic Growth," *Carnegie-Rochester Conference Series on Public Policy*, 1994, *40*, 1–57.

_ **and** _ , "International Data on Educational Attainment: Updates and Implications," *Center for International Development (CID) at Harvard University*, 2000, *No. 042.*

_ **and X. Sala-i-Martin**, *Economic Growth*, MIT Press, 2003.

Bates, J. and C. Granger, "The Combination of Forecasts," *Operational Research Quarterly*, 1969, *20*, 451–468.

Benhabib, J. and M. Spiegel, "The Role of Financial Development in Growth and Investment," *Journal of Economic Growth*, 2000, *5*, 341–360.

Berger, J., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

Bhargava, A. and J. D. Sargan, "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods," *Econometrica*, 1983, *51*, 1635–1659.

Blundell, R. and S. Bond, "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 2000, *19*, 321–340.

_ **and S.R. Bond**, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 1998, *87*, 115–143.

Bond, S.R., A. Hoeffler, and J. Temple, "GMM Estimation of Empirical Growth Models," *CEPR Discussion Papers*, 2001, *No. 3048.*

Brock, W. and S. Durlauf, "Growth Empirics and Reality," *World Bank Economic Review*, 2001, *15*, 229–272.

Buckland, S., K. Burnham, and N. Augustin, "Model Selection: An Integral Part of Inference," *Biometrics*, 1997, *53*, 603–618.

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

Chamberlain, G., "Panel Data," in Zvi Griliches and Michael Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland, 1984, pp. 1247–1318.

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

Ciccone, A. and M. Jarocinski, "Determinants of Economic Growth: Will Data Tell?," *forthcoming American Economic Journal: Macroeconomics*, 2009.

Claeskens, G. and N. Hjort, "The Focused Information Criterion," *Journal of the American Statistical Association*, 2003, *98*, 900–916.

_ and _ , *Model Selection and Model Averaging*, Cambirdge University Press, 2008.

Clemen, R., "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 1989, *5*, 559–583.

Cohen-Cole, E., S. Durlauf, J. Fagan, and D. Nagin, "Model Uncertainty and the Deterrent Effect of Capital Punishment," *Federal Reserve Bank of Boston Working Paper*, 2007.

Cremers, K., "Stock Return Predictability: A Bayesian Model Selection Perspective," *The Review of Financial Studies*, 2002, *15*, 1223–1249.

DeLong, J. and L. Summers, "Equipment Investment and Economic Growth," *Quarterly Journal of Economics*, 1991, *106*, 445–502.

Diebold, F. X. and J. A. Lopez, "Forecast Evaluation and Combination," *Handbook of Statistics*, 1996, pp. 241–268.

Durlauf, S., A. Kourtellos, and C. Tan, "Are Any Growth Theories Robust?," *Economic Journal*, 2008, *118*, 329–346.

_ , _ , and _ , "Is God in the Details? A Reexamination of the Role of Religion in Economic Growth," *Discussion Papers Series, Tufts University*, 2009, *No. 0613.*

_ , P. Johnson, and J. Temple, "Growth Econometrics," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1, Elsevier, 2005, chapter 8, pp. 555–677.

Easterly, W., "How Much Do Distortions Affect Growth?," *Journal of Monetary Economics*, 1993, *32*, 187–212.

Edgerton, H. and L. Kolbe, "The Method of Minimum Variation for the Combination of Criteria," *Psychometrika*, 1936, *1*, 183–188.

Eicher, T., A. Lenkoski, and A. Raftery, "Bayesian Model Averaging and Endogeneity Under Model Uncertainty: An Application to Development Determinants," 2009, *University of Washington Working Paper UWEC-2009-19.*

_ , C. Papageorgiou, and A. Raftery, "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants," *Journal of Applied Econometrics*, 2009, *forthcoming.*

Fernández, C., E. Ley, and M. Steel, "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, 2001, *100*, 381–427.

_ , _ , and _ , "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 2001, *16*, 563–576.

_ , _ , and _ , "Bayesian Modeling of Catch in a Northwest Atlantic Fishery," *Journal of the Royal Statistical Society, Series C*, 2002, *51*, 257–280.

Foster, D. and E. George, "The Risk Inflation Criterion for Multiple Regression," *The Annals of Statistics*, 1994, *22*, 1947–1975.

Frankel, J. and D. Romer, "Does Trade Cause Growth?," *American Economic Review*, 1999, *89*, 379–399.

Furnival, G. and R. Wilson, "Regression by Leaps and Bounds," *Technometrics*, 1974, *16*, 499–511.

Galbraith, J. and D. Hodgson, "Dimension Reduction and Model Averaging for Estimation of Artists' Age-Valuation Profiles," *CIRANO Working Paper*, 2009.

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

_ and J. Sachs, "The Economic Burden of Malaria," *American Journal of Tropical Medicine and Hygiene*, 2001, *64*, 85–96.

Garratt, A., K. Lee, H. Pesaran, and Y. Shin, "Forecast Uncertainties in Macroeconomic Modeling: An Application to the U.K. Economy," *Journal of the American Statistical Association*, 2003, *98*, 829–838.

Geisser, S., "A Bayes Approach for Combining Correlated Estimates," *Journal of the American Statistical Association*, 1965, *60*, 602–607.

George, E., "Discussion of "Model Averaging and Model Search Strategies" by M. Clyde," in J. Bernardo, A. Berger, P. Dawid, and Smith A., eds., *Bayesian Statistics*, Oxford University Press, 1999.

Halperin, M., "Almost Linearly-Optimum Combination of Unbiased Estimates," *Journal of the American Statistical Association*, 1961, *56*, 36–43.

Hansen, B., "Least Squares Model Averaging," *Econometrica*, 2007, *75*, 1175–1189.

_ and J. Racine, "Jackknife Model Averaging," *Unpublished Manuscript*, 2009.

Hansen, L. P., J. Heaton, and A. Yaron, "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 1996, *14*, 262–280.

Hausman, J. A. and W. E. Taylor, "Panel Data and Unobservable Individual Effects," *Econometrica*, 1981, *49*, 1377–1398.

Hjort, N. and G. Claeskens, "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 2003, *98*, 879–899.

Hoeting, J., D. Madigan, A. Raftery, and T. Volinsky, "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 1999, *14*, 382–417.

_ , _ , _ , and Volinsky C., "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 1999, *14*, 382–417.

Holtz-Eakin, D., W. Newey, and H. S. Rosen, "Estimating vector autoregressions with panel data," *Econometrica*, 1988, *56*, 1371–1395.

Horst, P., "Obtaining a Composite Measure from a Number of Different Measures of the same Attribute," *Psychometrika*, 1938, *1*, 53–60.

Islam, N., "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, 1995, *110* (4), 1127–1170.

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

Kass, R. and L. Wasserman, "A Reference Bayesian Test for Nested Hypothesis with Large Samples," *Journal of the American Statistical Association*, 1995, *90*, 928–934.

Kleibergen, F. and E. Zivot, "Bayesian and Classical Approaches to Instrumental Variable Regression," *Journal of Econometrics*, 2003, *114*, 29–72.

Knight, M., N. Loayza, and D. Villanueva, "Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach," *IMF Staff Papers*, 1992, *40*, 512–541.

Koop, G., *Bayesian Econometrics*, Wiley-Interscience, 2003.

Krugman, P., "Increasing Returns and Economic Geography," *Journal of Political Economy*, 1991, *99*, 483–499.

Kuersteiner, G. and R. Okui, "Constructing Optimal Instruments by First-Stage Prediction Averaging," *Econometrica*, 2010, *78*, 697–718.

Laplace, P. S., *Deuxime Supplément a la Théorie Analytique des Probabilités*, Courcier, Paris, 1818.

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

_ , "Let's Take the Con Out of Econometrics," *American Economic Review*, 1983, *73*, 31–43.

_ **and H. Leonard**, "Reporting the Fragility of Regression Estimates," *Review of Economics and Statistics*, 1983, *65*, 306–317.

Levine, R. and D. Renelt, "A sensivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 1992, *82*, 942–963.

Ley, E. and M. Steel, "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression," *Journal of Applied Econometrics*, 2009, *24*, 651–674.

Lucas, R., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 1988, *22*, 3–42.

Madigan, D. and A. Raftery, "Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam's Window," *Journal of the American Statistical Association*, 1994, *89*, 1535–1546.

_ **and J. York**, "Bayesian Graphical Models for Discrete Data," *International Statistical Review*, 1995, *63*, 215–232.

Magnus, J., O. Powell, and P. Prüfer, "A Comparison of Two Model Averaging Techniques with an Application to Growth Empirics," *Journal of Econometrics*, 2010, *154*, 139–153.

Mankiw, N. G., D. Romer, and D. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 1992, *107*, 407–437.

McAleer, M., A. Pagan, and P. Volker, "What Will Take the Con Out of Econometrics?," *American Economic Review*, 1985, *75*, 293–307.

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

Moral-Benito, E., "Determinants of Economic Growth: A Bayesian Panel Data Approach," *World Bank Policy Research Working Paper*, 2009, *No. 4830.*

_ , "Panel Growth Regressions with General Predetermined Variables: Likelihood-Based Estimation and Bayesian Averaging," *Unpublished Manuscript*, 2009.

Mundlak, Y., "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 1978, *46*, 69–85.

Nickell, S., "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 1981, *49*, 1417–1426.

Pesaran, H., C. Schleicher, and P. Zaffaroni, "Model Averaging in Risk Management with an Application to Futures Markets," *Journal of Empirical Finance*, 2009, *16*, 280–305.

Raftery, A., "Bayesian Model Selection in Social Research," *Sociological Methodology*, 1995, *25*, 111–163.

Rodríguez, F. and D. Rodrik, "Trade Policy and Economic Growth: A Skeptic's Guide to Cross-National Evidence," *NBER Working Papers*, 2000, *No. 7081.*

Romer, P., "Growth Based on Increasing Returns due to Specialization," *American Economic Review*, 1987, *77*, 56–62.

_ , "Endogenous Technological Change," *Journal of Political Economy*, 1990, *98*, 71–102.

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

Sala-i-Martin, X., "I Just Ran Two Million Regressions," *The American Economic Review*, 1997, *87*, 178–183.

_ , G. Doppelhofer, and R. Miller, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 2004, *94*, 813–835.

Solow, R., "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 1956, *70*, 65–94.

Stigler, S., "Laplace, Fisher, and the Discovery of the Concept of Sufficiency," *Biometrika*, 1973, *60*, 439–445.

Stock, J. and M. Watson, "Forecasting With Many Predictors," in G. Elliot, C. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Amsterdam: North-Holland, 2006, pp. 515–554.

Swan, T., "Economic Growth and Capital Accumulation," *Economic Record*, 1956, *32*, 334–361.

Tavares, J. and R. Wacziarg, "How Democracy Affects Growth," *European Economic Review*, 2001, *45*, 1341–1378.

Temple, J., "Robustness Tests of the Augmented Solow Model," *Journal of Applied Econometrics*, 1998, *13*, 361–375.

Timmermann, A., "Forecast Combinations," in G. Elliott, C. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, Amsterdam: North-Holland, 2006, pp. 135–196.

Tobias, J. and M. Li, "Returns to Schooling and Bayesian Model Averaging: A Union of Two Literatures," *Journal of Economic Surveys*, 2004, *18*, 153–180.

Volinsky, C., D. Madigan, A. Raftery, and R. Kronmal, “Bayesian Model Averaging in Proportional Hazard Models: Predicting the Risk of a Stroke,” *Applied Statistics*, 1997, *46*, 443–448.

Wagner, M. and J. Hlouskova, “Growth Regressions, Principal Components and Frequentist Model Averaging,” *Working Paper, Institute for Advanced Studies, Vienna*, 2009.

Wan, A. and X. Zhang, “On the Use of Model Averaging in Tourism Research,” *Annals of Tourism Research*, 2009, *36*, 525–532.

Wright, J., “Bayesian Model Averaging and Exchange Rate Forecasts,” *Journal of Econometrics*, 2008, *146*, 329–341.

_ , “Forecasting US inflation by Bayesian Model Averaging,” *Journal of Forecasting*, 2008, *28*, 131–144.

Zellner, A., “On Assessing Prior Distributions and Bayesian Regression Analysis With g-Prior Distributions,” in P. Goel and A. Zellner, eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Amsterdam: North-Holland/Elsevier, 1986.