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

Enrique Moral-Benito*
CEMFI

October 2009
JOB MARKET PAPER

ABSTRACT

In this paper I estimate empirical growth models simultaneously considering endogenous regressors and model uncertainty. In order to apply Bayesian methods such as Bayesian Model Averaging (BMA) to dynamic panel data models with predetermined or endogenous variables and fixed effects, I propose a likelihood function for such models. The resulting maximum likelihood estimator can be interpreted as the LIML counterpart of GMM estimators. Via Monte Carlo simulations, I conclude that the finite sample performance of the proposed estimator is better than that of the commonly used standard GMM. In contrast to previous consensus in the empirical growth literature, once endogeneity and model uncertainty are accounted for, empirical results indicate that the estimated convergence rate is not significantly different from zero. Moreover, there seems to be only one variable, the investment ratio, that robustly causes long-run economic growth.

JEL Codes: C11, C33, O40
Keywords: Dynamic Panel Estimation, Growth Regressions, Bayesian Model Averaging, Weak Instruments, Maximum Likelihood

---

*E-mail address: enrique.moral@gmail.com. I wish to thank Manuel Arellano for his constant guidance, advice and patience. I am also grateful to comments from Dante Amengual and seminar participants at CEMFI, the Mondragone-La Pietra-Moncalieri Doctoral Workshop at the Collegio Carlo Alberto, the Conference on Panel Data in Bonn and the EEA-ESEM Meetings in Barcelona. All programs used in this paper are available from my website.

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 paper 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 paper 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 generalized method of moments -henceforth 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 analysis 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 paper 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 to select 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. Fernandez et al. (2001) 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.

[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 545)

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 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.[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 paper 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

[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).

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-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 paper 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 paper, 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 paper. 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 paper 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 paper is strongly preferred to standard GMM estimators in terms of finite sample performance.

[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

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 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 Fernandez et al. (2001)).

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 can not be rejected given the available data. On the other hand, I conclude that there is only one variable that seems to robustly 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 paper is organized as follows. Section 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. In Section 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 5. Finally, Section 6 concludes and auxiliary results are gathered in the Appendix.

2 Dynamic Panel Data with Feedback: Likelihood-Based Estimation

Consider the following panel data model:

$$y_{it} = \alpha y_{it-1} + x'_{it}\beta + w'_i\delta + \eta_i + \zeta_t + v_{it} \tag{1}$$

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, w_i, \eta_i\right) = 0 \qquad (t = 1, ..., T)(i = 1, ..., N) \tag{2}$$

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 (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 (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 (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 (1)-(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.

2.1 Completing the General Predetermined Variables 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:

$$\begin{aligned} y_{i0} &= w_i'\delta_y + c_y\eta_i + v_{i0} && \text{(3a)} \\ x_{i1} &= \Delta_1 w_i + \gamma_{10} y_{i0} + c_1\eta_i + u_{i1} && \text{(3b)} \\ y_{i1} &= \alpha y_{i0} + x_{i1}'\beta + w_i'\delta + \eta_i + v_{i1} && \text{(3c)} \end{aligned}$$

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

$$\begin{aligned} x_{it} &= \Delta_t w_i + \gamma_{t0} y_{i0} + ... + \gamma_{t,t-1} y_{i,t-1} + \Lambda_{t1} x_{i1} + ... + \Lambda_{t,t-1} x_{i,t-1} + c_t\eta_i + u_{it} && \text{(3d)} \\ y_{it} &= \alpha y_{i,t-1} + x_{it}'\beta + w_i'\delta + \eta_i + v_{it} && \text{(3e)} \end{aligned}$$

Remark: Note that by writing the system as in (3a)-(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 (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:

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

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:

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

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:

$$\begin{aligned} L &= -\frac{N}{2} \ln \det \left(B^{-1} D \Omega D' B'^{-1}\right) \\ &- \frac{1}{2} tr \left\{ \left(B^{-1} D \Omega D' B'^{-1}\right)^{-1} \left[R - W(B^{-1}C)'\right]' \left[R - W(B^{-1}C)'\right] \right\} \end{aligned} \tag{4}$$

where R and X_t are the following matrices:

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

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 A.1 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 A.2, drastically reduces the number of paramaters to be estimated.

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.

3.1 Model and Estimators

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

$$y_{it} = \alpha y_{it-1} + \beta_1 x^1_{it-1} + \beta_2 x^2_{it-1} + \eta_i + v_{it} \tag{5}$$

$$E\left(v_{it} \mid y_{it-1}, ..., y_{i0}, x^1_{it-1}, ..., x^1_{i0}, x^2_{it-1}, ..., x^2_{i0}, \eta_i\right) = 0 \tag{6}$$

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 paper. 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 Appendix A.3.

[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).

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 (6) implies a set of linear moment conditions of the form:

$$E[w_i^{t-2}(\Delta y_{it} - \alpha \Delta y_{it-1} - \beta_1 \Delta x_{it-1}^1 - \beta_2 \Delta x_{it-1}^2)] = 0 \tag{7}$$

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 paper. 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 (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.

3.2 Results

Table 1 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 (5)-(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

Table 1: 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.

this paper 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 1 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 1, 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 1. 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 Appendix A.3. These additional results confirm the patterns that emerge from Table 1. 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 paper 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.

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:

$$\gamma_i = \beta \ln y_{i0} + \psi X_i + \epsilon_i \tag{8}$$

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 (8). An important objective of the present paper 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 (8) that can be called the canonical panel growth regression:

$$\ln y_{i,t} = (1+\beta)\ln y_{i,t-1} + \psi X_{i,t-1} + \eta_i + \zeta_t + v_{i,t} \qquad (i=1,...,N)(t=1,...,T) \tag{9}$$

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 (1983). 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

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

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 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:

$$y_{i,t} = \alpha y_{i,t-1} + \psi x_{i,t-1} + \eta_i + \zeta_t + v_{i,t} \tag{10a}$$

$$E\left(v_{i,t} \mid y_i^{t-1}, x_i^{t-1}, \eta_i\right) = 0 \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{10b}$$

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.

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 (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 $g+d$. Appendix A.4 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 2. 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 paper is applied to the basic Solow-Swan model and the results are shown in columns (4) and (8) of Table 2. 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.

[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.

Table 2: 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 (10a)-(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 2. Standard errors are in parenthesis.

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 paper 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.

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 (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 (1994) as well as an alternative specification explained below.

The basic empirical framework of Barro regressions with panel data is given by equation (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 (1994), among the 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 (1994). 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 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

[9] A more detailed description of the data sources and variables is in Appendix A.4

Table 3: 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 (1994) 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 (10a)-(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 2. Standard errors are in parenthesis.

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 (1994). 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 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.

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, Fernandez et al. (2001) 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 (2009) 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,

[10] See also Raftery (1995).

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.

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 paper 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 (2007), 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.

[11] A more detailed description of the data and its sources can be found in Appendix A.4

- 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 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 paper 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.

5.2 Bayesian Averaging of Maximum Likelihood Estimates (BAMLE)

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:

$$\Psi = \theta X + \epsilon \tag{11}$$

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

$$E(\theta|y) = \sum_{j=1}^{2^K} P\left(M_j|y\right) E(\theta|y, M_j) = \sum_{j=1}^{2^K} P\left(M_j|y\right) \widehat{\theta}^j_{ML} \tag{12}$$

where $\widehat{\theta}^j_{ML}$ is the ML estimate for model j. In this particular case, the sub-system LIML estimator presented in Section 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:

$$\begin{aligned} V(\theta|y) &= \sum\nolimits_{j=1}^{2^K} P\left(M_j|y\right) V(\theta|y, M_j) \\ &+ \sum\nolimits_{j=1}^{2^K} P\left(M_j|y\right) \left(E(\theta|y, M_j) - E(\theta|y)\right)^2 \end{aligned} \tag{13}$$

Inspection of (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 4) times the Posterior Inclusion Probability (PIP) in column 5 of Table 4. Similarly, the unconditional posterior variance can be computed according to $V(\theta|y)_{uncond} = [V(\theta|y)_{cond} + E^2(\theta|y)_{cond}] \times PIP - E^2(\theta|y)_{uncond}$.

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

[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.

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:

$$P(M_j|y) = \frac{P(M_j)(NT)^{\frac{-k_j}{2}} f(y|\widehat{\theta}_j, M_j)}{\sum_{i=1}^{2^K} P(M_i)(NT)^{\frac{-k_i}{2}} f(y|\widehat{\theta}_i, M_i)} \tag{14}$$

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 (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. (2009) conclude that this UIP combined with the uniform model prior (i.e. all models are equally probable *a priori*) I consider in the paper 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 4. 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 4 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 can not 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)

[14]Unweighted counterparts of the three measures in equations (12)-(13) are not reported here but they are available upon request.

[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 (2009), where model uncertainty and the endogeneity of the lagged dependent variable were considered.

Table 4: BAMLE Results

	Posterior mean conditional on inclusion (1)	Posterior s.d. conditional on inclusion (2)	Sign certainty probability (3)	Fraction of models with $\|tstat\| > 2$ (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 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 (12)). Column (2) refers to the square root of the posterior variance presented in equation (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.

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 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 4 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), Fernandez et al. (2001) and Moral-Benito (2009)), 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 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 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

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

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 1.

Figure 1: Posterior Distributions of Selected Coefficients

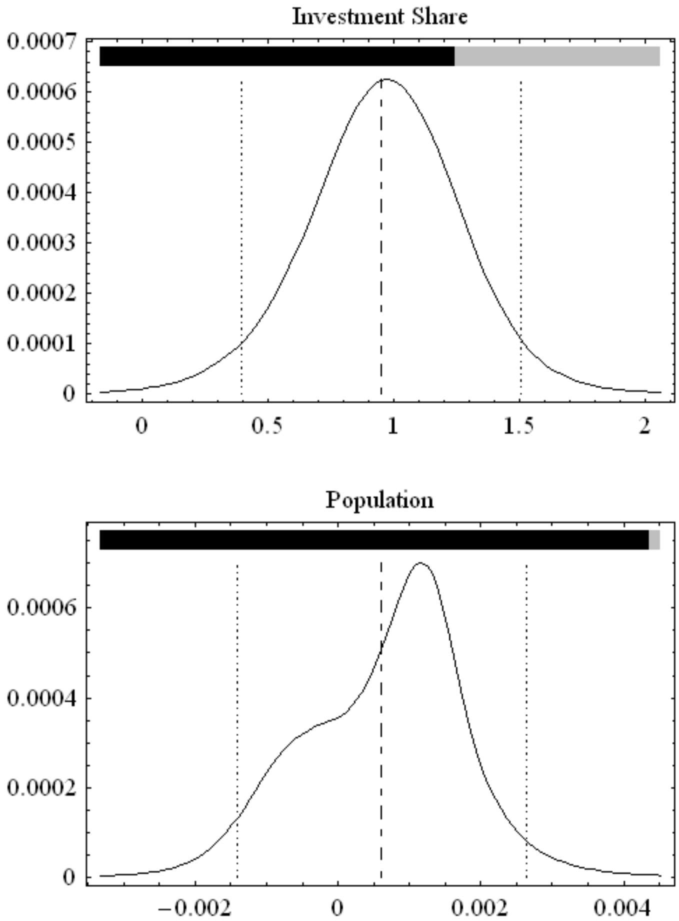

Figure 1 presents the marginal posterior distributions of the investment share and population coefficients. In particular, each graph consists of two parts: a gauge on top of the graphs that indicates the Posterior Inclusion Probability (PIP) of the variables and the normal mixture density for each coefficient. A dashed vertical line indicates the posterior mean conditional on inclusion presented in column 1 of Table 4. The equivalent to a classical 95% confidence interval is represented by two vertical dotted lines.

Analyzing Figure 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.

6 Concluding Remarks

This paper 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 paper is the first one 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 can not be rejected (at least accross the countries in my sample (see Appendix A.4)). 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 paper, 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.

A Appendix

A.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:

$$\eta_i = \gamma_0 y_{i0} + x'_{i1}\gamma_1 + \epsilon_i \tag{15}$$

Note that, again, in (15) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$ in order to ensure identification of δ.

Moreover, by substituing (15) in (1) the whole model can be written as follows:

$$y_{i1} = (\alpha + \gamma_0) y_{i0} + x'_{i1}(\beta + \gamma_1) + w'_i\delta + \epsilon_i + v_{i1} \tag{16a}$$

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

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + x'_{it}\beta + \gamma_0 y_{i0} + x'_{i1}\gamma_1 + w'_i\delta + \epsilon_i + v_{it} && (16b) \\ x_{it} &= \pi_{t0} y_{i0} + \pi_{t1} x_{i1} + \pi^w_t w_i + \xi_{it} && (16c) \end{aligned}$$

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:

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

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

$$B^S R^S_i = \Pi z_i + U_i \tag{17}$$

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:

$$\left(\begin{array}{c:c} B_{11}^S & B_{12}^S \\ \hdashline 0 & I_{k-1} \end{array}\right) \left(\begin{array}{c} R_{i1}^S \\ \hdashline R_{i2}^S \end{array}\right) = \left(\begin{array}{c} \Pi_1 \\ \hdashline \Pi_2 \end{array}\right) z_i + \left(\begin{array}{c} U_{i1} \\ \hdashline U_{i2} \end{array}\right) \tag{18}$$

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:

$$\Omega^S = Var(U_i) = Var\left(\begin{array}{c} U_{i1} \\ \hdashline U_{i2} \end{array}\right) = \left(\begin{array}{c:c} \Omega_{11}^S & \Omega_{12}^S \\ \hdashline \Omega_{21}^S & \Omega_{22}^S \end{array}\right) \tag{19}$$

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:

$$cov(\epsilon_i, \xi_{it}) = \phi_t \qquad \forall t = 2, ..., T \tag{20a}$$

$$cov(v_{ih}, \xi_{it}) = \begin{cases} \psi_{h,t} & \text{if } h < t \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{20b}$$

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]:

$$L_S \propto -\frac{N}{2} \ln \det(\Omega^S) - \frac{1}{2} tr\left((\Omega^S)^{-1} U'U\right) \tag{21}$$

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 $z_i = (y_{i0}, x_{i1}', w_i)'$:

$$f(y_i^T, x_i^T | z_i) = \int \prod_{t=1}^{T} f(y_{it} | y_i^{t-1}, x_i^t, w_i, \eta_i) \prod_{t=2}^{T} f(x_{it} | y_i^{t-1}, x_i^{t-1}, w_i, \eta_i) dG(\eta_i | z_i) \tag{22}$$

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 (21) is the same as in (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 contemporaneous relationships between the x variables and between the variables and the strictly exogenous variables). See Appendix A.2 for more details on the concentration of the SEM log-likelihood.

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

A.2 Concentrated Likelihood using the SEM Parametrization

Maximizing the log-likelihood in (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 (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 (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, (21) can be written as follows:

$$\begin{aligned} L &\propto -\frac{N}{2} \ln \det \Omega_{11} + \frac{N}{2} \ln \det G_{22} - \frac{1}{2} tr(\Omega_{11}^{-1} U_1' U_1) - tr(F_{12} G_{22} U_2' U_1) \\ &- \frac{1}{2} tr(G_{22} U_2' U_2) - \frac{1}{2} tr(F_{12} G_{22} F_{21} U_1' U_1) \end{aligned} \tag{23}$$

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

$$\begin{aligned}L_2 &\propto -\frac{N}{2}\ln\det\Omega_{11} + \frac{N}{2}\ln\det G_{22} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1)\\ &- \frac{1}{2}tr\{(R_2+U_1F_{12})'Q(R_2+U_1F_{12})G_{22}\}\end{aligned} \tag{24}$$

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:

$$L_3 \propto -\frac{N}{2}\ln\det\Omega_{11} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) - \frac{N}{2}\ln\det(\frac{1}{N}H) \tag{25}$$

A.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:

$$\begin{aligned} w_{it} &= (y_{it}, x^1_{it}, x^2_{it})' \\ \zeta_i &= (\zeta^y_i, \zeta^1_i, \zeta^2_i)' \\ \vartheta_{it} &= (\epsilon^y_{it}, \epsilon^1_{it}, \epsilon^2_{it})' \\ Var\left((w'_{i0}, \zeta'_i)'\right) &= \Omega_{MC} \\ Var\left(\vartheta_{it}\right) &= \Sigma_{MC} \end{aligned}$$

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:

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

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

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

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 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 A1: 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.

A.4 Data Appendix

Table A2: 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 A3: List of Countries

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			

References

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

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

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

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

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

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

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

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

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

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

——— (1996): "Democracy and Growth," *Journal of Economic Growth*, 1, 1–27.

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

——— (2000): "International Data on Educational Attainment: Updates and Implications," *CID Working Paper*, No. 042.

Barro, R. and X. Sala-i-Martin (2003): *Economic Growth*, MIT Press.

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

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

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

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

——— (2000): "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 19, 321–340.

Bond, S. R., A. Hoeffler, and J. Temple (2001): "GMM Estimation of Empirical Growth Models," *CEPR Discussion Papers*, 3048.

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

Chamberlain, G. (1983): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches and M. Intriligator, Amsterdam: North-Holland, 1247–1318.

Ciccone, A. and M. Jarocinski (2007): "Determinants of Economic Growth: Will Data Tell?" *CEPR Discussion Paper No. 6544*.

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

Durlauf, S. N., P. A. Johnson, and J. R. W. Temple (2005): "Growth Econometrics," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1, chap. 8, 555–677.

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

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

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

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

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

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

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

Hoeting, J., D. Madigan, A. Raftery, and V. C. (1999): "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–417.

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

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

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

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

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

Koop, G. (2003): *Bayesian Econometrics*, Wiley-Interscience.

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

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

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

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

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

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

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

——— (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98, 71–102.

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

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

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

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

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