SUPPLEMENTARY APPENDIX

Dynamic Panels with Predetermined Regressors: Likelihood-based Estimation and Bayesian Averaging with an Application to Cross-Country Growth

Enrique Moral-Benito

This appendix contains a version of the likelihood function for static panels with general predetermined regressors (Section A); some evidence on the finite sample performance of the sub-system LIML estimator when the Data Generating Process is not gaussian (Section B); a brief description of the variables typically considered as growth determinants (Section C); an overview of the literature on growth regressions based on single-model approaches with some models re-estimated employing the sub-system LIML estimator (Section D); a summary of the Bayesian model averaging methodology (Section E).

A Static Panels with Predetermined Regressors

Static panels with individual-specific effects and partially endogenous regressors are also of interest in practice. One prominent example is the estimation of production functions in which we typically face two problems: (i) the regressors (employment and stock of capital) are potentially correlated with firm-specific fixed effects and past 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 a good candidate to address the weak-instruments problem present in the estimation of production functions. By the same token, there are many other situations in which the econometric issues just described are also present.

In this Appendix we present the likelihood function for such a model. In particular, given the likelihood concentration procedure described in the paper based on the Simultaneous Equations Model (SEM) parametrization, we discuss here this representation for a static panel data model with partially endogenous regressors and fixed effects.

Let us consider a static panel data model as follows:

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

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

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: $x_i^t = (x'_{i1}, \ldots, x'_{it})'$. In the remaining of the exposition, we assume that all the variables are in deviations from their cross-sectional mean in order to rule out the common factors ζ_t. Assumption (2) accommodates partially endogenous regressors (xs) correlated with the fixed effects (ηs), and also strictly exogenous regressors ws.

Analogously to the dynamic case discussed in the paper, we can rewrite the model in matrix form as follows:

$$B^S R_i^S = \Pi z_i + U_i \tag{3}$$

where R_i^S and U_i are the vectors of data and errors defined in the paper.

The differences in this static version of the model arise in the coefficient matrices B^S and Π, and the $(k+m) \times 1$ vector of strictly exogenous variables z_i given now by:

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

The new matrices of structural coefficients B^S and reduced form coefficients Π are as follows:

$$B^S = \left(\begin{array}{c|c} I_T & B_{12}^S \\ \hline 0 & I_{k-1} \end{array}\right)$$

$$\Pi = \left(\begin{array}{c} \Pi_1 \\ \hline \Pi_2 \end{array}\right)$$

where:

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

$$\Pi_1 = \begin{pmatrix} \beta' + \gamma'_1 & \delta' \\ \gamma'_1 & \delta' \\ \vdots & \vdots \\ \gamma'_1 & \delta' \end{pmatrix}_{T \times (k+m)} \qquad \Pi_2 = \begin{pmatrix} \pi_{21} & \pi_2^w \\ \vdots & \vdots \\ \pi_{T1} & \pi_T^w \end{pmatrix}_{k(T-1) \times (k+m)}$$

Given the new matrices of coefficients (together with the normality assumption) the log-likelihood for the static model is analogous to the dynamic case:

$$L_S \propto -\frac{N}{2}\ln\det(\Omega^S) - \frac{1}{2}tr\left((\Omega^S)^{-1}U'U\right) \tag{4}$$

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.

The maximizer of L_S is a consistent and asymptotically normal estimator regardless of non-normality. In particular, the resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to one-step first-differenced GMM estimators discussed in Arellano and Bond (1991). In contrast to the dynamic case, note that assumption (2) does not imply lack of autocorrelation in the errors so that additional GMM moment conditions (e.g. Ahn and Schmidt (1995)) are not necessary for the asymptotic equivalence. However, the likelihood concentration procedure presented in the paper for the dynamic case is also valid in this static setting.

B Monte Carlo under Non-Normality

As discussed in the paper, neither the asymptotic distribution nor the finite sample behavior of the sub-system LIML estimator proposed in this paper are affected by the normality assumption. The reason is that in the linear case, the log-likelihood resulting from the normality assumption can be interpreted as a GMM objective function under a particular choice of weighting matrix (see Arellano (2003) pp.71-73). However, in order to illustrate that the finite sample performance of the estimator remains the same under non gaussian errors, this section presents some additional Monte Carlo results in which the data generating process is not normally distributed.

In the Monte Carlo results presented in the paper, the true Data Generating Process (DGP) is always normally distributed and thus the performance of the Gaussian LIML estimator introduced in this paper could be driven by this assumption. In Table A1 we can see the results of different Monte Carlo exercises in which the true DGP is non-normal. Six different non-normal cases are considered.

In the first three panels we present the Monte Carlo results under DGPs with tail behavior different from the normal case. In particular, in Panel A the DGP distribution is a mixture of normals with excess kurtosis $\kappa = 0.23$ instead of the 0 excess kurtosis of the normal distribution.[1] In order to further explore the robustness of the results with respect to different excess kurtosis in the true DGP distribution, in Panel B we simulate the data according to a mixture of two normals with a higher excess kurtosis ($\kappa = 1.99$). Finally, in Panel C, we assume that the true DGP is distributed as a t-student with 4 degrees of freedom that implies an infinite kurtosis so

[1]See Mardia (1970) for more details on the generation of multivariate mixtures of normal distributions with different excess kurtosis.

Table A1: Monte Carlo Results under non-normality

	$\alpha = 0.95$			$\beta_1 = 0.20$			$\beta_2 = -0.10$		
	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML	WG	diff GMM	sub-sys LIML
	Panel A: Mixture of normals $\kappa = 0.23$ and $\delta = 0$								
median	.428	.433	.893	.082	$-.110$	.143	$-.095$	$-.152$	$-.096$
iqr	.081	.323	.164	.107	.259	.194	.092	.157	.157
MAE	.522	.517	.077	.118	.312	.106	.047	.085	.078
	Panel B: Mixture of normals $\kappa = 1.99$ and $\delta = 0$								
median	.423	.437	.892	.084	$-.102$	.143	$-.097$	$-.145$	$-.096$
iqr	.090	.308	.167	.115	.262	.195	.100	.167	.156
MAE	.527	.513	.074	.116	.303	.108	.050	.088	.078
	Panel C: t-student with 4 degrees of freedom								
median	.433	.443	.901	.079	$-.119$	.141	$-.096$	$-.151$	$-.103$
iqr	.092	.303	.160	.111	.263	.190	.103	.164	.157
MAE	.517	.507	.069	.121	.319	.107	.052	.087	.079
	Panel D: Mixture of normals $\kappa = 0$ and $\delta = 5$								
median	.401	.648	.910	.190	.051	.159	$-.234$	$-.253$	$-.129$
iqr	.077	.234	.140	.125	.306	.321	.092	.172	.148
MAE	.549	.303	.066	.064	.185	.160	.134	.157	.077
	Panel E: Mixture of normals $\kappa = 0$ and $\delta = 50$								
median	.382	.661	.910	.232	.136	.196	$-.260$	$-.253$	$-.124$
iqr	.077	.241	.137	.132	.374	.457	.093	.168	.153
MAE	.568	.289	.069	.069	.190	.227	.160	.156	.076
	Panel F: Mixture of normals $\kappa = 1.99$ and $\delta = 50$								
median	.383	.656	.906	.232	.138	.187	$-.265$	$-.249$	$-.123$
iqr	.090	.257	.129	.140	.364	.434	.093	.174	.150
MAE	.567	.295	.067	.074	.192	.215	.165	.158	.077

Notes: 1,000 replications. iqr is the 75th-25th interquartile range; MAE denotes the median absolute error. κ indicates the excess of kurtosis, being $\kappa = 0$ the one corresponding to the normal distribution. δ refers to the difference in means of the normal distributions in the mixture. If $\delta = 0$ the mixture distribution is symmetric, otherwise is non-symmetric. Parameter values calibrated to ten-year periods data. In all panels the sample size is $T = 4, N = 100$.

that the tails of this distribution are much more thicker that the tails of a normal distribution. In all the three cases the results are virtually the same as in the normal case presented in Table 1 of the paper.

We depart from symmetric distributions in panels D and E. In Panel D, we simulate the true DGP according to a mixture of normals with 0 excess of kurtosis ($\kappa = 0$) as the normal distribution but with a non-symmetric shape. More concretely, we use a mixture of two different normal distributions with different means ($\delta = 5$ indicates that the difference between both means is 5) so that the resulting distribution is non-symmetric. An alternative non-symmetric distribution is considered in Panel E in which the difference between the means is larger, $\delta = 50$. The results remain practically unchanged in both non-symmetric cases.

Finally, in order to explore the robustness of the results to non-symmetric distributions with thicker than normal tails, we consider in Panel E a mixture of normals with excess kurtosis $\kappa = 1.99$ and difference in means $\delta = 50$. This means that we are departing from the normal distributions in two aspects at the same time, we have a distribution which is clearly non-symmetric and has much thicker tails that the normal distribution. The Monte Carlo results show the same conclusion, the sub-system LIML estimator presented in the paper is strongly preferred to diff-GMM under errors that are far from normal.

C Growth Determinants

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 (2010) and Moral-Benito (2011), the fewer the potential growth determinants considered, the smaller the sensivity of the results.

In the paper we consider the following candidate growth determinants:

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

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

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

D Growth Empirics: Revisiting the Evidence

The bulk of the growth empirics literature is based on single model regressions (e.g. Barro (1991); Islam (1995); Caselli et al. (1996)). In this section we put the sub-system LIML estimator discussed in the paper at work in comparison with other commonly-used estimators in the "single model" growth regressions industry. The aim is twofold: on the one hand we revisit the evidence on the Solow model and Barro regressions estimates; and, on the other hand, we check the differences which arise between alternative estimators.

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:

$$\Upsilon_i = c \ln y_{i0} + \beta X_i + \epsilon_i \tag{5}$$

where $\Upsilon_i = t^{-1}(\ln y_{it} - \ln y_{i0})$ represents the growth rate of output per worker between 0 and t. On the other hand, X_i is a vector of variables that might include 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.

There is an important variant of the baseline empirical growth regression in (5) that can be called the canonical panel growth regression:

$$\ln y_{it} = (1+c)\ln y_{it-1} + \beta X_{it} + \eta_i + \zeta_t + v_{it} \qquad (i=1,...,N)(t=1,...,T) \tag{6}$$

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_{it} 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_{it}.[2] However, in these studies 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_{it}. 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 plausible in the growth context. Finally, Caselli et al. (1996) and Benhabib and Spiegel (2000) among others, take into consideration the predetermined nature[3] of the x variables allowing for the mentioned feedback

[2] This point refers to the fact that, by construction, all leads of y_{it-1} are correlated with v_{it} and, therefore, the within-groups estimator will produce biased estimates in the typical small-T growth panel. In order to address this issue, these studies employ the Π-matrix method discussed in Chamberlain (1984).

[3] This predetermined nature is also labeled as partial endogeneity in the paper, and it is sometimes denominated

process. In particular, in order to estimate the model, Caselli et al. (1996) and Benhabib and Spiegel (2000) use generalized method of moments (GMM) following techniques advanced by Holtz-Eakin et al. (1988) and Arellano and Bond (1991). The assumption that the explanatory variables are predetermined implies a set of moment restrictions that can be used in the context of GMM to generate consistent and efficient estimates of the parameters of interest. More concretely, the employed moment restrictions can be interpreted as an instrumental variables model where lagged levels of the variables are used as instruments for their first-differences. As Blundell and Bond (1998) pointed out, with persistent series such as GDP, lagged levels may be only weak instruments for the equation in first-differences. Thus, in spite of being consistent as N goes to infinity, this estimator is poorly behaved in finite samples. For this reason, these GMM estimates 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.

The likelihood-based estimator presented in the paper 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, LIML estimators might 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.

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

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \zeta_t + v_{it} \tag{7a}$$

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, \eta_i\right) = 0 \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{7b}$$

where $\alpha = 1 + c$, $y_{i,t}$ is the GDP per capita for country i in period t, x_{it} 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_{it} 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 we aim to avoid the effect of business-cycle fluctuations and, therefore, focus on the long-term growth process. However, we also present some estimations using 5-year periods data.

weakly exogeneity in the growth literature.

D.1 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_{it} in (7a) 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 our 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, we choose 0.05 as a reasonable assessment of the value of $g + d$.

Table A2: Solow-Swan Model Estimation Results

	Five-year data ($T = 8$)				Ten-year data ($T = 4$)			
	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)	(0.037)	(0.014)	(0.050)	(0.112)	(0.091)
$\ln(s_{i,t-1})$	0.088	0.091	0.035	0.095	0.167	0.166	0.009	0.222
	(0.010)	(0.018)	(0.034)	(0.025)	(0.019)	(0.036)	(0.085)	(0.066)
$\ln(n_{i,t-1}+g+d)$	−0.204	−0.137	0.128	0.020	−0.441	−0.327	0.557	−0.102
	(0.041)	(0.071)	(0.108)	(0.100)	(0.085)	(0.163)	(0.325)	(0.309)
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 (7a)-(7b), where both fixed effects and (partial) endogeneity 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 discussed in the paper. Standard errors are in parenthesis. Replication material can be found in http://www.moralbenito.com.

We 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 A2. 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 an OLS variant where given the availability of a panel dataset, country dummies can be included in order to allow for the presence of unobserved heterogeneity correlated with the regressors (i.e. country-specific fixed effects). The results when employing both OLS and WG estimators are in line with previous literature. Columns (3) and (6) report first-differenced GMM estimates in the spirit of Caselli et al. (1996). The similarity between WG and diff-GMM estimates of the convergence parameter 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 accounting for potential endogeneity of the regressors, the diff-GMM estimates might not be 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[4] and the results are shown in columns (4) and (8) of Table A2. Inspection of these columns points to the importance of the finite-sample biases in previous first-differenced GMM estimates of this model. In contrast to previous panel estimates of the rate of convergence using the Solow-Swan framework, we 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[5] is roughly zero in both cases. However, the 95% confidence intervals are consistent with convergence rates that vary from -1.7% to 1.2% in the case of five-year periods data and from -2.0% to 1.5% 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 is reinforced using alternative specifications in this section, and in the paper when model uncertainty is also taken into consideration.

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.

D.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_{it} in (7a). In this subsection we apply

[4]A STATA command called **xtmoralb** that implements this estimator is available from my website http://www.moralbenito.com

[5]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.

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, we 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 (7a). Two kind of variables are included in theses regressions, first, initial levels of state variables measured at the beginning of the period (we 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 we 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, we 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 our sample period, the other two control variables are slightly different from those employed in the original specification but they capture similar effects. We 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.

Table A3 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 -1.3% to 1.8%. On the other hand, the conclusions with respect to other explanatory variables may change 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.

In columns (5)-(8) we present the results from an alternative specification. Imagine a researcher who is testing the effect of democracy on growth. For this purpose, she estimates a growth regression using as state variables the initial level of per capita GDP, the average years of secondary education and the country's population (in millions of people), and as a control variable she decides to only include the domestic investment ratio (I/GDP). Given this specification, the sub-system LIML 95% confidence interval for the convergence rate estimate goes from -0.9% to 3.0%. However, diff GMM provides convergence rate estimates in the range from 3.5% to 12.9% which might be upward biased due to weak instruments. Moreover, while the effect of investment (I/GDP) is estimated to be not significantly different from zero according to diff GMM, it is much larger in magnitude and significant according to sub-sys LIML. We thus conclude that the

Table A3: Barro Regressions Estimation Results

	Baseline Specification Ten-year data ($T = 4$)				Alternative Specification Ten-year data ($T = 4$)			
	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)	(0.077)	(0.019)	(0.051)	(0.107)	(0.088)
Education	0.040	0.039	0.055	0.030	0.016	0.036	0.076	0.030
	(0.015)	(0.036)	(0.081)	(0.066)	(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)	(0.356)				
I/GDP	0.588	0.781	0.857	1.114	0.891	0.797	0.351	1.268
	(0.133)	(0.213)	(0.279)	(0.298)	(0.132)	(0.193)	(0.284)	(0.321)
G/GDP	−0.246	−0.465	−0.314	−0.546				
	(0.115)	(0.284)	(0.534)	(0.496)				
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)	(0.087)	(0.042)	(0.058)	(0.082)	(0.095)
Population					0.0003	0.0017	0.002	0.0012
					(0.0001)	(0.0003)	(0.0003)	(0.0003)
Implied λ	0.017	0.038	0.017	0.002	0.003	0.047	0.082	0.011
	(0.003)	(0.008)	(0.009)	(0.008)	(0.002)	(0.008)	(0.024)	(0.010)
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 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 (7a)-(7b) where both fixed effects and (partial) endogeneity 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 discussed in the paper. Standard errors are in parenthesis. Replication material can be found in http://www.moralbenito.com.

consideration of the estimator discussed in the paper might be of interest for empirical growth researchers as an alternative to first-differenced GMM.

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. 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 might lead us to misleading conclusions even if we consider appropriate estimation techniques for a given model because we can not be sure whether this is the correct empirical model or not. To some extent, this fact might serve as an illustration of the need to take into account model uncertainty in empirical growth regressions.

E Bayesian Model Averaging

This section presents a brief overview of the BMA techniques considered to obtain the empirical results in Table 2 of the paper. Formally, consider a generic representation of an empirical model of the form:

$$\Psi = \theta X + \upsilon \tag{8}$$

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 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.[6] To be more precise, the point estimate of interest will be the mean of the posterior distribution of the parameters given the data:

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

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:[7]

$$E(\theta|data) = \sum_{j=1}^{2^K} P\left(M_j|data\right) E(\theta|data, M_j) = \sum_{j=1}^{2^K} P\left(M_j|data\right) \widehat{\theta}^j_{ML} \tag{9}$$

[6]A more detailed discussion of the BMA methodology can be found in Hoeting et al. (1999), Koop (2003) or Moral-Benito (2010) among others.

[7]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.

where $\widehat{\theta}^j_{ML}$ is the ML estimate for model j. In particular, we can consider the sub-system LIML estimator discussed in the paper or any other likelihood-based estimator emerging from a proper likelihood function.

Given the endogenous regressors setting considered in the paper, each of the models being considered here comprise the same set of simultaneous equations (i.e. each model is given by a set of structural form equations for the dependent variable in each time period and the same set of reduced form equations for the endogenous regressors). Therefore, model-specific sub-system LIML estimators maximize the joint density of the dependent variable and all the partially endogenous regressors. In order to guarantee comparability of the likelihoods, this is so even when some of the regressors are not "included" in the model, i.e. a given regressor is excluded from a particular model by simply restricting to zero its coefficient in the structural form equation. However, the key issue is that all the reduced form equations comprise the full set of endogenous regressors and thus are the same for all the models under consideration. By doing so, the densities of the different models are comparable.[8]

Similarly to the posterior mean, following Leamer (1978) we can also compute the posterior variance:

$$\begin{aligned} V(\theta|data) &= \sum\nolimits_{j=1}^{2^K} P\left(M_j|data\right) V(\theta|data, M_j) \\ &+ \sum\nolimits_{j=1}^{2^K} P\left(M_j|data\right) \left(E(\theta|data, M_j) - E(\theta|data)\right)^2 \end{aligned} \tag{10}$$

Inspection of (10) shows that the variance incorporates both the estimated variances of the individual models as well as the variance in estimates of the coefficients across different models. Hence, the uncertainty at the two different levels mentioned in the main text 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.[9]

Moreover, in this paper we consider model weights (i.e. the posterior model probabilities $P\left(M_j|data\right)$) based on the Schwarz asymptotic approximation to the Bayes Factor, and therefore:

$$P\left(M_j|data\right) = \frac{P\left(M_j\right)\left(NT\right)^{\frac{-k_j}{2}} f(data|\widehat{\theta}_j, M_j)}{\sum_{i=1}^{2^K} P\left(M_i\right)\left(NT\right)^{\frac{-k_i}{2}} f(data|\widehat{\theta}_i, M_i)} \tag{11}$$

where $f(data|\widehat{\theta}_j, M_j)$ is the maximized likelihood function for model j. Kass and Wasserman (1995) show that the Schwarz asymptotic approximation formula in (11) could also be obtained

[8]See Moral-Benito (2010) for more details on the combination of Bayesian model averaging and endogenous regressors.

[9]This means that when computing both of them from the posterior distribution we only consider the models in which the coefficient of the regressor in the structural equation is not restricted to be zero. However, the unconditional posterior mean can be easily obtained by multiplying the conditional posterior mean (column (1) in Table 2) times the Posterior Inclusion Probability (PIP) in column 5 of Table 2. Similarly, the unconditional posterior variance can be computed according to $V(\theta|data)_{uncond} = [V(\theta|data)_{cond} + E^2(\theta|data)_{cond}] \times PIP - E^2(\theta|data)_{uncond}$.

with a reasonable prior on the parameter space[10] 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 so that $P(M_j) = 1/2^K \forall j$) we 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.

Finally, BMA also considers the posterior probability (PIP) that a particular variable h is included in the regression. In particular, 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. The PIP of variable h is calculated as the sum of the posterior model probabilities for all of the models including that particular variable:

$$\text{PIP} = P(\theta_h \neq 0|data) = \sum\nolimits_{\theta_h \neq 0} P(M_j|data). \qquad (12)$$

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.

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.

[10] 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.

Barro, R. and J. 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," *Center for International Development (CID) at Harvard University*, No. 042.

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

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., A. Hoeffler, and J. Temple (2001): "GMM Estimation of Empirical Growth Models," *CEPR Discussion Papers*, No. 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. (1984): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches and M. Intriligator, Amsterdam: North-Holland, 1247–1318.

Ciccone, A. and M. Jarocinski (2010): "Determinants of Economic Growth: Will Data Tell?" *American Economic Journal: Macroeconomics*, 2(4), 223–247.

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

Durlauf, S., P. Johnson, and J. 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.

Frankel, J. and D. Romer (1999): "Does Trade Cause Growth?" *The 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.

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," *The 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.

Mardia, K. (1970): "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, 36, 519–530.

Moral-Benito, E. (2010): "Model Averaging in Economics," *CEMFI Working Paper*, 10/08.

——— (2011): "Determinants of Economic Growth: A Bayesian Panel Data Approach," *forthcoming The Review of Economics and Statistics.*

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

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

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

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.