DYNAMIC PANELS WITH PREDETERMINED REGRESSORS: LIKELIHOOD-BASED ESTIMATION AND BAYESIAN AVERAGING WITH AN APPLICATION TO CROSS-COUNTRY GROWTH

2011

Enrique Moral-Benito

Documentos de Trabajo
N.º 1109

BANCO DE ESPAÑA
Eurosistema

DYNAMIC PANELS WITH PREDETERMINED REGRESSORS: LIKELIHOOD-BASED ESTIMATION AND BAYESIAN AVERAGING WITH AN APPLICATION TO CROSS-COUNTRY GROWTH

DYNAMIC PANELS WITH PREDETERMINED REGRESSORS: LIKELIHOOD-BASED ESTIMATION AND BAYESIAN AVERAGING WITH AN APPLICATION TO CROSS-COUNTRY GROWTH

Enrique Moral Benito (*)

BANCO DE ESPAÑA

(*) All programs and data used in this paper together with replication instructions are available from my website http://www.moralbenito.com. In this web site you can also find the STATA command xtmoralb which implements the estimator proposed in the paper. A previous version circulated under the title «Panel Growth Regressions with General Predetermined Variables: Likelihood-Based Estimation and Bayesian Averaging». Contact: Bank of Spain. Alcalá 48, 28014, Madrid. enrique.moral@gmail.com. I wish to thank Manuel Arellano for his constant guidance and advice. I am also grateful to comments from Dante Amengual, Steve Bond, Stéphane Bonhomme, Antonio Ciccone, Bruce Hansen, Marek Jarocinski, Eduardo Ley, Oliver Linton, Michael Manove, Claudio Michelacci, Hashem Pesaran, Adrian Raftery, Enrique Sentana, and seminar participants at CEMFI, University of Warwick, University of St. Gallen, UAB, GRIPS, Bank of Spain, the Mondragone-LaPietra-Moncalieri Doctoral Workshop at the Collegio Carlo Alberto, the Conference on Panel Data in Bonn, the EEA Meetings in Barcelona, the European Winter Meeting of the Econometric Society in Budapest, and the Econometric Society World Congress in Shanghai.

Documentos de Trabajo. N.º 1109

2011

ISSN: 0213-2710 (print)
ISSN: 1579-8666 (on line)
Depósito legal: M. 21861-2011
Unidad de Publicaciones, Banco de España

Abstract

This paper discusses likelihood-based estimation of linear panel data models with general predetermined variables and individual-specific effects. The resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to standard GMM but tends to have smaller finite-sample biases as illustrated in simulation experiments. Moreover, the availability of such a likelihood function allows applying the Bayesian apparatus to this class of panel data models. Combining the aforementioned estimator with Bayesian model averaging methods we estimate empirical growth models simultaneously considering endogenous regressors and model uncertainty. Empirical results indicate that only the investment ratio seems to robustly cause long-run economic growth. Moreover, the estimated rate of convergence is not significantly different from zero.

Keywords: Dynamic panel estimation, maximum likelihood, weak instruments, growth regressions, bayesian model averaging**.**

JEL classification: C11, C33, O40.

Resumen

En este documento se analiza la estimación por máxima verosimilitud de modelos lineales de datos de panel con efectos fijos y regresores endógenos. El estimador máximo verosímil resultante es asintóticamente equivalente a estimadores de panel por el Método Generalizado de Momentos (Arellano y Bond, 1991) pero tiene menores sesgos en muestras finitas como se ilustra en las simulaciones. Por otra parte, la disponibilidad de una función de verosimilitud permite aplicar métodos Bayesianos a esta clase de modelos de datos de panel. En concreto, combinando el estimador propuesto con métodos Bayesianos de promediado de modelos se estiman ecuaciones de crecimiento atajando simultáneamente los problemas de endogeneidad e incertidumbre del modelo. Los resultados empíricos obtenidos indican que sólo la inversión parece ser causante robusto del crecimiento económico a largo plazo. Por otra parte, la tasa de convergencia estimada no es significativamente diferente de cero.

Palabras claves: Datos de panel, Máxima verosimilitud, Instrumentos débiles, Regresiones de crecimiento, Promediado Bayesiano de modelos**.**

Códigos JEL: C11, C33, O40.

1 Introduction

In this paper we consider a linear (dynamic) panel data model with general predetermined explanatory variables and unobservable individual effects. Such a model is typically estimated by panel IV techniques like first-differenced GMM, e.g. Holtz-Eakin et al. (1988), Arellano and Bond (1991). However, in practice the application of GMM often entails finite sample biases, especially when the instruments are weak (i.e. lagged levels of the variables are weakly correlated with subsequent first-differences). A number of alternative methods have been considered to address this issue from a method-of-moments perspective (e.g. Hansen et al. (1996); Alonso-Borrego and Arellano (1999); Arellano and Bover (1995)). In contrast, in this paper we focus on likelihood-based estimation of this class of models. The aim is twofold: on the one hand, the likelihood counterpart of first-differenced GMM estimators is expected to alleviate finite sample biases due to weak instruments; on the other hand, the availability of such a likelihood function allows applying Bayesian methods such as Bayesian model averaging to panel data models with general predetermined variables.

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). Thus, method-of-moments estimators (like first-differenced GMM) exploit many overidentifying restrictions, although the quality of the instruments is often poor. In order to consider the LIML counterpart for this kind of panel IV estimators, we construct the likelihood function of a dynamic panel data model with general predetermined variables and individual effects correlated with the regressors. Hansen et al. (1996) and Akashi and Kunitomo (2010) among others have also considered LIML estimators for such a panel model. However, these are only LIML analog estimators in the sense of the minimax instrumental-variable interpretation given by Sargan (1958) to the original LIML estimator; therefore they do not provide suitable likelihood functions.

Proper likelihood-based approaches for dynamic panel models with unobservable individual effects have been discussed in the literature (e.g. Bhargava and Sargan (1983); Alvarez and Arellano (2003)). The focus in these approaches is on the distribution of the dependent variable conditional on a set of exgoneous regressors. In this paper we construct the joint likelihood function of the dependent variable and a set of predetermined (or partially endogenous) regressors conditional on the initial observations, and optionally, on additional exogenous variables. Intuitively, we complete the model with an unrestricted feedback process which is specified in the form of period-specific linear projections of the non-exogenous variables on all available lags. Moreover, the analysis is marginal on the individual effects which can be correlated with the regressors.

The resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to one-step first-differenced GMM augmented with moments implied by the serial correlation properties of the errors.[1] Simulation experiments serve to evaluate the finite-sample behavior of the proposed estimator. Our simulation results show that the estimator has negligible biases in contrast to the commonly-used Arellano and Bond's (1991) GMM estimator, which has large biases, especially when the generated series are persistent over time. Therefore, we conclude that the proposed likelihood-based estimator is preferred to standard GMM estimators in terms of finite-sample performance.

Researchers interested in "not large N, small T" panels might often face this weak-instruments problem. Panel growth regressions are probably the best example: 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. Moreover, 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 commonly-used first-differenced GMM estimator is poorly behaved in the growth framework (e.g. Bond et al. (2001)). The likelihood-based estimator discussed in this paper provides a promising alternative.

Model uncertainty represents also a challenge to empirical growth researchers. It emerges because theory does not provide enough guidance to select the proper empirical model, and results in a total of more than 140 variables proposed as growth determinants (see for instance Durlauf et al. (2005)). One commonly-used alternative to address model uncertainty is Bayesian model averaging — henceforth BMA — methods which construct parameter estimates that formally address the dependence of model-specific estimates on a given model. Fernández et al. (2001) and Sala-i-Martin et al. (2004) popularized the use of BMA in the growth context under the assumption of exogenous growth determinants. In order to simultaneously address model uncertainty and different forms of endogeneity, the combination of BMA with IV and panel data models is an interesting line of open research (e.g. Moral-Benito (2011); Durlauf et al. (2008); Eicher et al. (2009a)). The availability of the suitable likelihood function derived in this paper allows us to combine BMA methods (or the Bayesian apparatus in general) with panel models

[1]The additional moments are quadratic restrictions of the type discussed in Ahn and Schmidt (1995). On the other hand, we refer here to fixed T, large N asymptotics.

under the assumption of endogenous regressors. The possibility to simultaneously address the problems of model uncertainty and endogeneity seems of paramount importance for empirical growth researchers.[2]

Empirical results cast doubt on previous consensus in the growth regressions literature. On the one hand, we do not find evidence of conditional convergence across the countries in the sample. In particular, the estimated speed of convergence is 0.73%, but it is not significantly different from zero. On the other hand, only the investment ratio can be labeled as a robust determinant of economic growth accordingly to the Bayesian robustness check used in the paper.

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 (i.e. predetermined regressors). Monte Carlo evidence on the finite-sample behavior of the estimator is provided in Section 3. Results from combining the estimator and model averaging techniques are presented in Section 4. Finally, Section 5 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 given the dynamics of the model is considered in assumption (2).[3] The model also relaxes the strict exogeneity assumption for the x variables that are also considered as predetermined (this is why we refer to the model as having general predetermined variables) allowing for feedback from lagged values of y to the current value for x. More precisely, assumption (2) implies that the x variables in period t are correlated with past shocks $(v_{i0}, ..., v_{it-1})$ but uncorrelated with present and future shocks $(v_{it}, ..., v_{iT})$. Other intermediate configurations can be accommodated in this framework. For instance, we might be interested in allowing for non-zero correlations between the partially endogenous regressors

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

[3]Assumption (2) also implies lack of autocorrelation in v_{it} since lagged vs are linear combinations of the variables in the conditioning set.

(x_{it}) and contemporaneous shocks (v_{it}). We label the x variables as predetermined or partially endogenous[4] as opposed to the other two possible configurations, namely, strict exogeneity (if x_{it} is uncorrelated with the full path of shocks $v_{i0}, ..., v_{iT}$) and strict endogeneity (if x_{it} is correlated with all the shocks from $t = 0$ to $t = T$). A static (i.e. without lagged dependent variable) version of this panel data model with partially endogenous regressors and its likelihood function are discussed in Appendix A.1.

The model also incorporates m strictly exogenous regressors that may or may not have temporal variation. In the remaining of the exposition we 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) we 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. The term ζ_t in (1) captures unobserved common factors across units in the panel and, therefore, this particular form of cross-sectional dependence is allowed.[5]

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); Alonso-Borrego and Arellano (1999)). In order to alleviate this problem, several alternatives have been proposed in the literature from a method-of-moments perspective (see for example Arellano and Bover (1995), Hansen et al. (1996), Blundell and Bond (1998), Alonso-Borrego and Arellano (1999) and Akashi and Kunitomo (2010)). The alternatives discussed in Hansen et al. (1996) and Akashi and Kunitomo (2010) are usually labeled as LIML approaches. However, they are method-of-moments estimators which can be interpreted as LIML analog estimators given the minimax instrumental-variable interpretation to the original LIML estimator discussed in Sargan (1958).

Given the available evidence in the single equation case, in this paper we adopt a likelihood-based perspective which is expected to be a good candidate in the face of the weak-instruments problem in this panel setting. Moreover, the availability of such a suitable likelihood function allows combining 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) or Alvarez and Arellano (2003)). Therefore, the focus was on the distribution of y_i^T conditional on the regressors and, sometimes on the initial observation y_{i0}. On the other hand, it is possible to either condition

[4]This configuration is sometimes denominated weakly exogeneity in the panel growth regressions literature.

[5]In practice, this is done by simply working with cross-sectional de-meaned data. In the remaining of the exposition, we assume that all the variables are in deviations from their cross-sectional mean.

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 the macro forecasting literature, the likelihood will be fundamentally misspecified. Here instead we specify the distribution of the regressors and present the proper 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 in (1) 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, we specify the feedback process in the form of cross-sectional linear projections of the partially endogenous x variables on all available lags, having period-specific coefficients. The complete model is therefore as follows:[6]

$$y_{i0} = w_i'\delta_y + c_y\eta_i + v_{i0} \tag{3a}$$

$$x_{i1} = \Delta_1 w_i + \gamma_{10} y_{i0} + c_1\eta_i + u_{i1} \tag{3b}$$

$$y_{i1} = \alpha y_{i0} + x_{i1}'\beta + w_i'\delta + \eta_i + v_{i1} \tag{3c}$$

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

$$x_{it} = \Delta_t w_i + \gamma_{t0} y_{i0} + ... + \gamma_{t,t-1} y_{i,t-1} + \Lambda_{t1} x_{i1} + ... + \Lambda_{t,t-1} x_{i,t-1} + c_t\eta_i + u_{it} \tag{3d}$$

$$y_{it} = \alpha y_{i,t-1} + x_{it}'\beta + w_i'\delta + \eta_i + v_{it} \tag{3e}$$

> **Remark:** Note that by writing the system as in (3a)-(3e) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$, since otherwise we 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:

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

[6] Note 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.

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, we 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, we 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, we are now able to write the system in matrix form as follows:

$$BR_i = Cw_i + D\Xi_i \tag{4}$$

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. Note that the block-diagonal variance-covariance matrix Ω allows for time-series heteroskedasticity.

Finally, under normal errors the log-likelihood of the model given by (4) can be written as:

$$\begin{aligned} L &= -\frac{N}{2}\ln\det\left(B^{-1}D\Omega D'B'^{-1}\right) \\ &- \frac{1}{2}tr\left\{\left(B^{-1}D\Omega D'B'^{-1}\right)^{-1}\left[R - W(B^{-1}C)'\right]'\left[R - W(B^{-1}C)'\right]\right\} \end{aligned} \quad (5)$$

where R and X_t are the following matrices:

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

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

It is important to remark here that the maximizer of L is a consistent and asymptotically normal estimator regardless of non-normality. In particular, the resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to standard GMM estimators because the resultant first order conditions correspond to a GMM problem with a convenient choice of weighting matrix (see Arellano (2003) pp.71-73). More specifically, it corresponds to the Arellano and Bond's (1991) GMM estimator augmented with the moments discussed in Ahn and Schmidt (1995) and employing the optimal weighting matrix under normality and conditional homoskedasticity.

This parametrization of the complete model is labeled as Full Covariance Structure (FCS) representation. In this parametrization, 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 represents a concern since the number of parameters to be estimated becomes intractable.

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, we present in the the next section 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 us to concentrate out all the parameters of the dynamic relationships between the xs which are not of central interest. This concentration (described in Appendix A.2) drastically reduces the number of parameters to be estimated so that the optimization problem becomes feasible and computationally affordable.

2.2 Simultaneous Equations Model (SEM) Representation

In this section we present the Simultaneous Equations Model (SEM) representation that allows us to concentrate some reduced form parameters of the resulting log-likelihood in order to make its maximization feasible and computationally affordable. 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.

We first define:

$$\eta_i = \gamma_0 y_{i0} + x'_{i1}\gamma_1 + \epsilon_i \tag{6}$$

Note that, again, in (6) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$ in order to ensure identification of δ.

Moreover, by substituing (6) in (1) the whole model can be written as follows:

$$y_{i1} = (\alpha + \gamma_0)y_{i0} + x'_{i1}(\beta + \gamma_1) + w'_i\delta + \epsilon_i + v_{i1} \tag{7a}$$

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

$$y_{it} = \alpha y_{i,t-1} + x'_{it}\beta + \gamma_0 y_{i0} + x'_{i1}\gamma_1 + w'_i\delta + \epsilon_i + v_{it} \tag{7b}$$

$$x_{it} = \pi_{t0}y_{i0} + \pi_{t1}x_{i1} + \pi^w_t w_i + \xi_{it} \tag{7c}$$

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, we 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 we are now able to rewrite the model in matrix form as follows:

$$B^S R^S_i = \Pi z_i + U_i \tag{8}$$

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 we additionally define the following vectors:

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

it is then possible to rewrite:

$$\left(\begin{array}{c:c} B^S_{11} & B^S_{12} \\ \hdashline 0 & I_{k-1} \end{array}\right)\left(\begin{array}{c} R^S_{i1} \\ \hdashline R^S_{i2} \end{array}\right) = \left(\begin{array}{c} \Pi_1 \\ \hdashline \Pi_2 \end{array}\right) z_i + \left(\begin{array}{c} U_{i1} \\ \hdashline U_{i2} \end{array}\right) \tag{9}$$

where:

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

$$\Pi_1 = \begin{pmatrix} \alpha+\gamma_0 & \beta'+\gamma'_1 & \delta' \\ \gamma_0 & \gamma'_1 & \delta' \\ \vdots & \vdots & \vdots \\ \gamma_0 & \gamma'_1 & \delta' \end{pmatrix}_{T\times(1+k+m)} \qquad \Pi_2 = \begin{pmatrix} \pi_{20} & \pi_{21} & \pi^w_2 \\ \vdots & \vdots & \vdots \\ \pi_{T0} & \pi_{T1} & \pi^w_T \end{pmatrix}_{k(T-1)\times(1+k+m)}$$

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:

$$\Omega^S = Var(U_i) = Var\left(\begin{array}{c} U_{i1} \\ \hdashline U_{i2} \end{array}\right) = \left(\begin{array}{c:c} \Omega^S_{11} & \Omega^S_{12} \\ \hdashline \Omega^S_{21} & \Omega^S_{22} \end{array}\right) \tag{10}$$

where:

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

$$\Omega^S_{11} = \sigma^2_\epsilon \iota\iota' + \begin{pmatrix} \sigma^2_{v_1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma^2_{v_T} \end{pmatrix}$$

  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 we can write:

$$cov(\epsilon_i, \xi_{it}) = \phi_t \qquad \forall t = 2, ..., T \tag{11a}$$

$$cov(v_{ih}, \xi_{it}) = \begin{cases} \psi_{h,t} & \text{if } h < t \\ \mathbf{0} & \text{otherwise} \end{cases} \tag{11b}$$

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

$$\Omega_{12}^S = \begin{pmatrix} \phi_2' + \psi_{1,2}' & \phi_3' + \psi_{1,3}' & \dots & \phi_T' + \psi_{1,T}' \\ \phi_2' & \phi_3' + \psi_{2,3}' & \dots & \phi_T' + \psi_{2,T}' \\ \phi_2' & \phi_3' & \dots & \phi_T' + \psi_{3,T}' \\ \vdots & \vdots & \ddots & \vdots \\ \phi_2' & \phi_3' & \dots & \phi_T' + \psi_{T-1,T}' \\ \phi_2' & \phi_3' & \dots & \phi_T' \end{pmatrix}_{T \times (T-1)k}$$

In view of matrix Ω_{12}^S and equations (11a)-(11b) is illustrative to describe how to accommodate other partial endogeneity configurations in addition to the baseline assumption presented in equation (2). For example, allowing for non-zero correlations between x_{it} and contemporaneous shocks (v_{it}) is straightforward by incorporating additional non-zero elements in the Ω_{12}^S matrix. More specifically, if we substitute assumption (2) by the alternative $E\left(v_{it} \mid y_i^{t-1}, x_i^{t-1}, w_i, \eta_i\right) = 0$ we shall substitute (11a)-(11b) by:

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

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

Under normal errors the log-likelihood for the model can be written as:[7]

$$L_S \propto -\frac{N}{2} \ln \det(\Omega^S) - \frac{1}{2} tr\left((\Omega^S)^{-1} U'U\right) \tag{12}$$

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

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

$$f(y_i^T, x_i^T | z_i) = \int \prod_{t=1}^{T} f(y_{it} | y_i^{t-1}, x_i^t, w_i, \eta_i) \prod_{t=2}^{T} f(x_{it} | y_i^{t-1}, x_i^{t-1}, w_i, \eta_i) dG(\eta_i | z_i) \tag{13}$$

As in the case of the FCS representation in the previous section, the maximizer of L_S is a consistent and asymptotically normal estimator regardless of non-normality. The resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to the Arellano and Bond's (1991) one-step GMM estimator augmented with the moments discussed in Ahn and Schmidt (1995). This is so because the (pseudo) likelihood function discussed here can be interpreted as the resulting GMM objective function when we combine the moment conditions in Arellano and Bond (1991) and Ahn and Schmidt (1995) using the optimal weighting matrix under normality and conditional homoskedasticity.[8]

Finally, the number of parameters to be estimated in (12) is the same as in the corresponding log-likelihood for the FCS parametrization (see equation (5)). This number might be intractable in practice and therefore, in order to make the problem feasible we consider the concentrated log-likelihood with respect to the unrestricted parameters in the matrices Π_2 and Ω_{22}^S (i.e. the parameters that capture the dynamic and contemporaneous relationships between the explanatory variables). See Appendix A.2 for more details on the concentration of the SEM log-likelihood.

3 Monte Carlo Simulation

In this section, we 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:

$$y_{it} = \alpha y_{it-1} + \beta_1 x_{it}^1 + \beta_2 x_{it}^2 + \eta_i + v_{it} \tag{14}$$

$$E\left(v_{it} \mid y_{it-1}, ..., y_{i0}, x_{it}^1, ..., x_{i1}^1, x_{it}^2, ..., x_{i1}^2, \eta_i\right) = 0 \tag{15}$$

[8] The first order conditions of the (pseudo) maximum likelihood estimator are true regardless of the normality and conditional homoskedasticity assumptions.

Suppose we have a random sample of individual time series of size T: $(\Theta'_{i1}, ..., \Theta'_{iT})'$ where $\Theta_{it} = (y_{it-1}, x^1_{it}, x^2_{it})'$ and $(i = 1, ..., N)$. On the other hand, we assume that initial observations $\Theta_{i1} = (y_{i0}, x^1_{i1}, x^2_{i1})'$ are observed. We further assume that the initial observations and the fixed effect are jointly normally distributed[9] 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.

Since empirical growth regressions is probably the most common situation in which general predetermined regressors arise, the baseline 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 we 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 the empirical exercises of this paper. On the other hand, since five-year periods are also commonly considered in empirical panel growth regressions, for the purpose of robustness, we 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.

Three alternative estimators are applied to the simulated samples. We 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 Θ 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 right-hand-side variables) 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, we consider the diff-GMM estimator commonly employed in panel growth regressions since Caselli et al. (1996). The assumption in equation (15) implies a set of linear moment conditions of the form:

$$E[\Theta_i^{t-1}(\Delta y_{it} - \alpha \Delta y_{it-1} - \beta_1 \Delta x^1_{it} - \beta_2 \Delta x^2_{it})] = 0 \tag{16}$$

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

[9]Note that the consistency of the estimators we consider in the Monte Carlo exercise is unaffected by the normality assumption (see Arellano (2003) pp.71-73). Moreover, in Appendix A.4 you can find additional Monte Carlo results under non-normality. These results illustrate that the finite sample behavior of the estimators remains the same under non-normality.

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 because it includes a set of structural-form equations and a set of reduced-form equations.

Under homoskedasticity, sub-system LIML is asymptotically equivalent to a GMM estimator that in addition to (16) 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 (14)-(15) (means and standard deviations are not reported because the sub-system LIML estimators can be expected to have infinite moments).

In the baseline specification in Panel A, N is fixed to 100 since it is the number of cross-section observations we find in a typical growth regression. On the other hand, given the main focus of this paper is on ten-year periods over the years 1960-2000, $T = 4$ is the number of available time series observations. This sample size in the within time dimension ($T = 4$) is also common in typical micro panels. 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

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.

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.

On the other hand, 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 conclusion 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.

Finally, Appendix A.4 presents additional Monte Carlo results under non-normality of the true Data Generating Process (DGP). Since the results remain virtually unchanged for distributional assumptions far from normal, we can conclude that the better finite sample performance of the sub-system LIML estimator is true regardless of the normality assumption in the Monte Carlo design.

4 Application to Cross-Country Growth

As pointed out by Durlauf et al. (2005), the stylized facts of economic growth have led to two major issues in the development of formal econometric analyses of growth. The first one revolves around the question of convergence: are contemporary differences in growth rates across countries transient over sufficiently long time horizons? The second issue concerns the identification of growth determinants: which factors seem to explain observed differences in aggregate economies? These two questions have been addressed by a huge literature on empirical growth regressions.

The canonical cross-country growth regression in its panel version takes the form:[10]

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \zeta_t + v_{it} \tag{17}$$

where y_{it} 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. Appendix A.6 provides an overview of the growth determinants we consider in this paper.

Problems with estimating such an empirical growth model are well known. The x variables are in general (partially) endogenous, and omitted variable bias arises due to the presence of country-specific effects (η_i) correlated with the regressors (assumption (2) summarizes this situation). In order to address these issues, first-differenced GMM estimators applied to dynamic panel data models has been commonly-used in empirical growth research. Given the persistence of series such as GDP or investment, these GMM estimators are expected to suffer from weak instruments so that the likelihood-based estimator discussed in this paper represents a promising alternative. In Appendix A.5 we provide a more detailed discussion about the estimation of empirical growth models as well as empirical evidence on the performance of competing estimators in this framework.

Another relevant challenge in the growth regressions literature is the issue of model uncertainty. This problem arises due to the lack of clear theoretical guidance on the choice of growth regressors to include in the vector x_{it} that results in a wide set of possible specifications. There-

[10] This specification corresponds to the model in equation (1) but without time-invariant regressors.

fore, 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 sensitivity of claims about the parameter of interest to model choice. Bayesian model averaging (BMA) represents an alternative to incorporate the uncertainty at the two levels described above.

The availability of the likelihood function discussed in Section 2 allows us to combine the resulting maximum likelihood estimator with BMA techniques in order to simultaneously address endogeneity and model uncertainty.

4.1 Model Averaging and Growth Empirics

A promising approach to account for model uncertainty is to employ Bayesian model averaging techniques to construct parameter estimates that formally address the dependence of model-specific estimates on a given model.[11] The fundamental principle of BMA is to treat both models and parameters as unobservable, and to estimate their distributions based on the observable data.[12] The basics of Bayesian model averaging are presented in Appendix A.7.

Sala-i-Martin et al. (2004) and Fernández et al. (2001) popularized the use of BMA in the growth regressions literature. More concretely, following techniques advanced by Raftery (1995), 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. In a pure Bayesian spirit, Fernández et al. (2001) consider alternative priors with the same objective. However, both studies rely on the exogeneity assumption of the regres-

[11] An alternative approach is based on model selection, i.e. the task of selecting a statistical model from a set of potential models given data. Given this approach, after the model selection step, both the inference and the conclusions of the analysis are typically based on the single model selected, and thus the uncertainty associated with the specification of the empirical model is somehow ignored. A good overview of this literature can be found in Claeskens and Hjort (2008).

[12] There also exists a frequentist approach to model averaging (e.g. Claeskens and Hjort (2003), Hansen (2007), Hansen and Racine (2010)); the main differences between frequentist and Bayesian model averaging arise from how model weights are selected and how inference is carried out. Compared with the frequentist approach, there has been an enormous literature on the use of BMA in statistics and more recently in economics. Thus, the BMA toolkit is larger than that of its frequentist counterpart.

sors and focus on cross-sectional data.[13] Moral-Benito (2011) extends the approach to a panel data setting simultaneously considering country-specific effects and partial endogeneity of the lagged dependent variable. In particular, Moral-Benito (2011) combines BMA with the likelihood function presented in Alvarez and Arellano (2003) for dynamic panels with exogenous regressors. Other studies such as Tsangarides (2004), Durlauf et al. (2008), Mirestean and Tsangarides (2009), Eicher et al. (2009a), and Durlauf et al. (2009) incorporate endogenous regressors and combine method-of-moments estimates with model averaging techniques.[14] In this section we combine the proper likelihood function previously introduced with BMA techniques in order to simultaneously address partial endogeneity of the regressors and model uncertainty in the context of (panel) growth empirics.

4.2 Empirical Results

Table 2 presents the results when combining the panel likelihood-based estimator presented in Section 2 with the Bayesian model averaging techniques described in Appendix A.7. In the context of empirical growth regressions, this combination represents an attempt to simultaneously consider model uncertainty and endogeneity of growth regressors (see Appendix A.6 for more details on the growth data considered in the paper).

Regarding the issue of convergence, the point estimate of the rate of convergence[15] of an economy to its steady state is 0.73%, much lower than previous panel studies such as Caselli et al. (1996) who estimated a convergence rate of around 12%. Moreover, the estimate of the rate of convergence is not significantly different from zero once we consider both levels of uncertainty described above (i.e. looking at the posterior s.d. resultant from the BMA approach). Therefore we cannot reject the null hypothesis of no conditional convergence across the countries in the sample.[16] This finding casts doubt on the conventional wisdom of conditional convergence as a strong empirical regularity in the country-level data (e.g. Barro and Sala-i Martin (1992), Caselli et al. (1996)).[17] For illustrative purposes we plot in Figure 1 the BMA posterior distribution of

[13] Magnus et al. (2010) and Masanjala and Papageorgiou (2008) also consider BMA methods in the framework of growth regressions with exogenous regressors.

[14] More specifically, these approaches consider pseudo likelihood functions, and hence the statistical justification of averaging method-of-moments estimates remains an open debate. Heuristically, these papers replace the fully specified likelihood by the adjusted method-of-moments objective function. Moral-Benito (2010) provides a more detailed discussion on the combination of model averaging with endogenous regressors.

[15] We estimate the rate of convergence as $\lambda = \frac{\ln \alpha}{-\tau}$ where $\tau = 10$ and α is the coefficient on $\ln(y_{t-1})$. Moreover, note that initial GDP ($\ln(y_{t-1})$) is included in all the models under consideration since theory offers strong guidance for this variable (see Durlauf et al. (2005)).

[16] This result was previously found in Moral-Benito (2011), where model uncertainty and the endogeneity of the lagged dependent variable were considered.

[17] For example, early versions of endogenous growth theories (e.g. Romer (1987, 1990) and Aghion and Howitt (1992)) were criticized because in contrast to the neoclassical growth model, they no longer predicted conditional

Table 2: BAMLE Results

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

Notes: In this table, the sub-system LIML estimator introduced in Section 2 is combined with the BMA methodology as described in Appendix A.7. 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 each particular variable. Column (2) refers to the square root of the posterior variance which incorporates model-specific uncertainty as well as uncertainty across alternative models. Column (3) presents the percentage of models in which the coefficient is significantly different from zero (either positive or negative). Finally, column (4) presents the Bayesian posterior inclusion probability of a given variable which 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. Replication material can be found in http://www.moralbenito.com.

the convergence coefficient which presents a substantial amount of probability mass on both sides of one.[18]

convergence.

[18]Analogously to the posterior mean, BMA posterior distributions are weighted averages of marginal posterior distributions conditional on each individual model. More concretely, these posteriors are mixture normal distributions because model-specific posteriors are normal. This is so because we make use of the Bernstein-von Mises theorem, also known as the Bayesian CLT (Berger (1985) provides an in-depth analysis and an excellent illustration.), 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 posterior distributions consist of two parts, a continuous distribution on the real line and a point mass at zero. Therefore, in addition to the continuous mixture normal distribution a gauge that represents the Posterior Inclusion Probability (PIP) of the variables is also plotted.

Figure 1: Posterior Distribution of the Convergence Coefficient

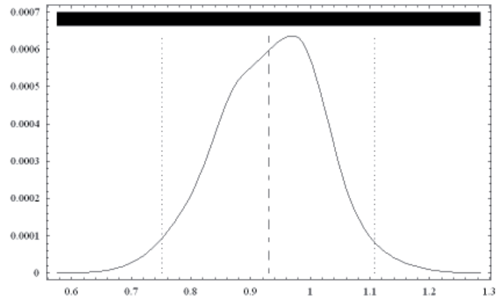

Figure 1 presents the marginal posterior distribution of the coefficient on the lagged dependent variable (i.e. the convergence coefficient). The graph consists of two parts: a gauge on top of the graph that indicates the Posterior Inclusion Probability (PIP) of the variable (which is 1 by definition since we include the lagged dependent variable in all the models under consideration) and the normal mixture density for the coefficient's posterior distribution. A dashed vertical line indicates the posterior mean conditional on inclusion presented in column 1 of Table 2. The equivalent to a classical 95% confidence interval is represented by two vertical dotted lines. Note that in this case, a coefficient equal to 1 means that there is no evidence of conditional convergence.

The empirical evidence on growth determinants seems to be conclusive for only one variable, the investment ratio. This is so because its posterior mean is three times its posterior standard deviation.[19] For the rest of the growth determinants, their corresponding posterior standard deviations are high enough to preclude them from having a significant effect on economic growth (note that these posterior variances incorporate not only the uncertainty conditional on a given model as usual, but also the uncertainty across different models). On the other hand, the investment ratio is only a proximate determinant of economic growth according to Rodrik (2003) and Acemoglu (2009). Indeed, to the extent that growth might be driven by other fundamental determinants (e.g. institutions), the causality may well run backwards despite our efforts to account for feedback effects in this paper.

For further insights we can see in Figure 2 the full BMA posterior distributions of the coefficients that correspond to the variables investment share and population. In particular, we observe that the estimated effect of investment on growth is unambiguously positive. 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, its marginal posterior distribution presents probability mass on both sides of zero, indicating that its effect on growth could be either positive or negative. As shown in Table 2, this is also the case for all the remaining candidate growth determinants considered

[19]While the ratio of posterior mean to posterior standard deviation is not distributed according to the usual t-distribution, Sala-i-Martin et al. (2004) note that in most cases, having a ratio around two in absolute value indicates an approximate 95-percent Bayesian coverage region that excludes zero. This 'pseudo-t' statistic indicates that in the case of the investment ratio, its positive effect on growth is significantly different from zero. Moreover, in 98.8% of the individual models its coefficient was estimated to be significant at the 95% level.

in the paper (see Appendix A.6).

This result is in contrast to previous findings in the literature.[20] In particular, previous BMA studies applied to growth regressions always find that several regressors (not necessarily coincident) are robustly related to economic growth (e.g. Sala-i-Martin et al. (2004), Fernández et al. (2001), Durlauf et al. (2008), Mirestean and Tsangarides (2009), Moral-Benito (2011)).

Two conclusions are drawn from this lack of robustness result; first, that the fragility of cross-country growth regressions is such that casts doubt on the validity of this approach to shed light on the issue of long-run growth determinants; secondly, that there may not be universal rules about what makes countries grow.

Figure 2: Posterior Distributions of Selected Coefficients

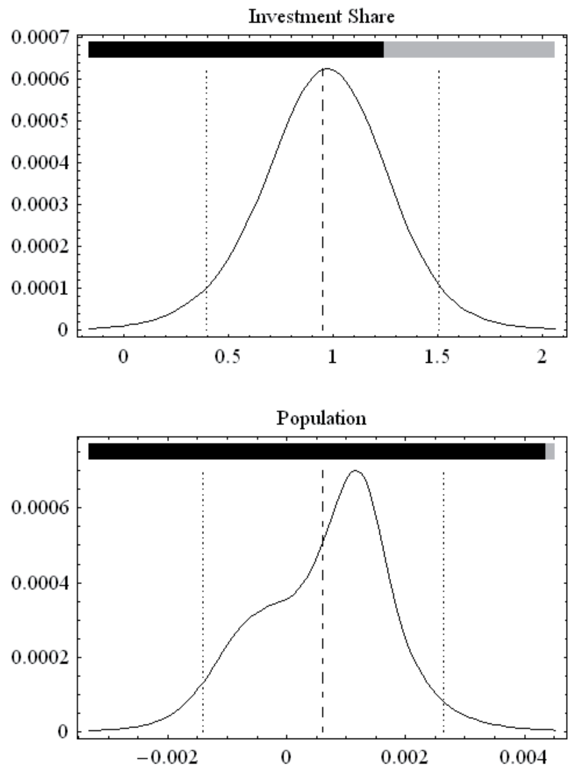

Figure 2 presents the marginal posterior distributions of the investment share and population coefficients. In particular, each graph consists of two parts: a gauge on top of the graphs that indicates the Posterior Inclusion Probability (PIP) of the variables and the normal mixture density for each coefficient's posterior distribution. A dashed vertical line indicates the posterior mean conditional on inclusion presented in column 1 of Table 2. The equivalent to a classical 95% confidence interval is represented by two vertical dotted lines.

[20] Note that single-model results considering a panel likelihood function with partially endogenous regressors but ignoring model uncertainty provide evidence in favor of several variables robustly related to economic growth (see Appendix A.5). On the other hand, BMA results considering model uncertainty and a panel likelihood function with exogenous regressors also provide evidence of a (different) set of variables robustly related to growth (see Moral-Benito (2011))

5 Concluding Remarks

In this paper we discuss likelihood-based estimation of a linear (dynamic) panel data model with general predetermined explanatory variables and unobservable individual effects. The resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to one-step first-differenced GMM augmented with moments implied by the serial correlation properties of the errors (e.g. Holtz-Eakin et al. (1988), Arellano and Bond (1991), Ahn and Schmidt (1995)). Since the application of first-differenced GMM often entails finite sample biases, especially when the instruments are weak, simulation experiments are conducted to evaluate the finite-sample behavior of competing estimators. The simulation results show that the proposed likelihood-based estimator has negligible biases in contrast to the commonly-used Arellano and Bond's (1991) GMM estimator, which has large biases, especially when the generated series are persistent over time. Therefore, we conclude that the proposed likelihood-based estimator is preferred to standard GMM estimators in terms of finite-sample performance. This result can be interpreted as a generalization of the single equation case (see for example Anderson et al. (1982)).

The availability of a proper likelihood function allows us to combine the aforementioned estimator with Bayesian model averaging methods (or the Bayesian apparatus in general) in order to simultaneously address endogeneity and model uncertainty in the context of growth regressions. Once both issues are accounted for, the empirical results indicate that the hypothesis of lack of conditional convergence cannot be rejected. 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. On the other hand, in contrast to previous consensus in the BMA and growth literature, only the investment ratio can be labeled as a robust determinant of economic growth accordingly to the Bayesian robustness check considered in the paper.

A Appendix

A.1 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 Appendix A.2 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:

$$y_{it} = x_{it}'\beta + w_i'\delta + \eta_i + \zeta_t + v_{it} \tag{18}$$

$$E\left(v_{it} \mid x_i^t, w_i, \eta_i\right) = 0 \qquad (t = 1, ..., T)(i = 1, ..., N) \tag{19}$$

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 (19) accommodates partially endogenous regressors (xs) correlated with the fixed effects (ηs), and also strictly exogenous regressors ws.

Analogously to the dynamic case discussed in Section 2.2, we can rewrite the model in matrix form as follows:

$$B^S R_i^S = \Pi z_i + U_i \tag{20}$$

where R_i^S and U_i are the vectors of data and errors defined in Section 2.2.

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:

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

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

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

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

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.

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 (19) 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 Appendix A.2 for the dynamic case is also valid in this static setting.

A.2 Concentrated Likelihood using the SEM Parametrization

Maximizing the log-likelihood in (12) may be cumbersome (or even impossible) since the dimension of the numerical optimization problem is enormous. In particular, the number of parameters to be estimated (p) in (12) 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 (12) will be concentrated with respect 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, we 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, (12) can be written as follows:

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

Note that we can also write $\Omega_{11}^{-1} = G_{11} - G_{12} G_{22}^{-1} G_{21}$ and we have added and subtracted 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 (22) 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:

$$\begin{aligned} \widehat{U}_2' U_1 &= R_2' Q U_1 - F_{21} U_1' M U_1 \\ \widehat{U}_2' \widehat{U}_2 &= R_2' Q R_2 + F_{21} U_1' M U_1 F_{12} \end{aligned}$$

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

$$\begin{aligned} L_2 &\propto -\frac{N}{2} \ln\det \Omega_{11} + \frac{N}{2} \ln\det G_{22} - \frac{1}{2} tr(\Omega_{11}^{-1} U_1' U_1) \\ &- \frac{1}{2} tr\{(R_2 + U_1 F_{12})' Q (R_2 + U_1 F_{12}) G_{22}\} \end{aligned} \tag{23}$$

Step 2: Concentrating out Ω_{22}

We now turn to the concentration of L_2 with relation to Ω_{22}. Note that the log-likelihood is now written in terms of G_{22} and therefore, in practice we 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:

$$H = (R_2 + U_1 F_{12})' Q (R_2 + U_1 F_{12})$$

Therefore:

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

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:

$$L_3 \propto -\frac{N}{2} \ln\det \Omega_{11} - \frac{1}{2} tr(\Omega_{11}^{-1} U_1' U_1) - \frac{N}{2} \ln\det(\frac{1}{N} H) \tag{24}$$

A.3 Monte Carlo Details

For simulating the data in the Monte Carlo experiment, we first estimate a tri-variate VAR process for GDP[21] (y), investment ratio (x^1) and population growth (x^2). In particular, we consider the following VAR process:

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

where:

$$\begin{aligned} \Theta_{it} &= (y_{it-1}, x_{it}^1, x_{it}^2)' \\ \zeta_i &= (\zeta_i^y, \zeta_i^1, \zeta_i^2)' \\ \vartheta_{it} &= (\epsilon_{it}^y, \epsilon_{it}^1, \epsilon_{it}^2)' \\ Var\left((\Theta_{i1}', \zeta_i')'\right) &= \Omega_{MC} \\ Var\left(\vartheta_{it}\right) &= \Sigma_{MC} \end{aligned}$$

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

1. Generate Θ_{i1} and ζ_i according to $(\Theta_{i1}', \zeta_i')' \sim N(0, \widehat{\Omega}_{MC})$.
2. For $t = 2, ..., T$:
   (a) Generate ϑ_{it} according to $\vartheta_{it} \sim N(0, \widehat{\Sigma}_{MC})$
   (b) Then generate Θ_{it} according to $\Theta_{it} = \widehat{\Gamma}\Theta_{it-1} + \zeta_i + \vartheta_{it}$

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

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

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

[21] 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 we 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 Monte Carlo under Non-Normality

As discussed in the main text, 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 Appendix presents some additional Monte Carlo results in which the data generating process is not normally distributed.

In the Monte Carlo results presented so far, the true Data Generating Process (DGP) was always normally distributed and thus the performance of the Gaussian LIML estimator introduced in this paper could be driven by this assumption. In the following Table 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.[22] 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 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.

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 dis-

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

tributions 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 this paper is strongly preferred to diff-GMM under errors that are far from normal.

Table A2: 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 $T = 4, N = 100$.

A.5 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 Appendix we put the sub-system LIML estimator discussed in this 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:

$$\Upsilon_i = c \ln y_{i0} + \beta X_i + \epsilon_i \tag{25}$$

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 (25) that can be called the canonical panel growth regression:

$$\ln y_{it} = (1+c) \ln y_{it-1} + \beta X_{it} + \eta_i + \zeta_t + v_{it} \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{26}$$

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}.[23] 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[24] of the x variables allowing for the mentioned feedback 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 main text of this 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 vari-

[23] 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).

[24] This predetermined nature is also labeled as partial endogeneity in the main text of this paper, and it is sometimes denominated weakly exogeneity in the growth literature.

able 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:

$$y_{it} = \alpha y_{it-1} + \beta x_{it} + \eta_i + \zeta_t + v_{it} \tag{27a}$$

$$E\left(v_{it} \mid y_i^{t-1}, x_i^t, \eta_i\right) = 0 \qquad (i = 1, ..., N)(t = 1, ..., T) \tag{27b}$$

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.

A.5.1 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 (27a) 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 $g + d$. Appendix A.6 contains more details about the employed data.

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 A3. 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

Table A3: 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 (27a)-(27b), 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 presented in Section 2. Standard errors are in parenthesis. Replication material can be found in http://www.moralbenito.com.

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[25] and the results are shown in columns (4) and (8) of Table A3. 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[26] is roughly zero in both cases.

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

[26] 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.

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 Appendix, and in the main text 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.

A.5.2 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 (27a). In this subsection we 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, 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 (27a). 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.[27]

[27]A more detailed description of the data sources and variables can be found in Appendix A.6.

Table A4: 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 (27a)-(27b) 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 presented in Section 2. Standard errors are in parenthesis. Replication material can be found in http://www.moralbenito.com.

Table A4 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 consideration of the estimator discussed in this 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.

A.6 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 this paper we consider the following candidate growth determinants:[28]

- 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

[28]Table A5 presents more details on these variables and their sources.

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.

Table A5: 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 A6: 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			

A.7 Bayesian Model Averaging

This Appendix presents a brief overview of the BMA techniques considered to obtain the empirical results in Section 4 of the main text. Formally, consider a generic representation of an empirical model of the form:

$$\Psi = \theta X + \upsilon \tag{28}$$

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.[29] 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:[30]

$$E(\theta|data) = \sum_{j=1}^{2^K} P\left(M_j|data\right) E(\theta|data, M_j) = \sum_{j=1}^{2^K} P\left(M_j|data\right) \widehat{\theta}^j_{ML} \tag{29}$$

where $\widehat{\theta}^j_{ML}$ is the ML estimate for model j. In particular, we can consider the sub-system LIML estimator presented in Section 2 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

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

[30] 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.

set of reduced form equations for the endogenous regressors). Therefore, model-specific subsystem 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.[31]

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

$$\begin{aligned} V(\theta|data) &= \sum\nolimits_{j=1}^{2^K} P\left(M_j|data\right) V(\theta|data, M_j) \\ &+ \sum\nolimits_{j=1}^{2^K} P\left(M_j|data\right) \left(E(\theta|data, M_j) - E(\theta|data)\right)^2 \end{aligned} \tag{30}$$

Inspection of (30) 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.[32]

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:

$$P\left(M_j|data\right) = \frac{P\left(M_j\right)\left(NT\right)^{\frac{-k_j}{2}} f(data|\widehat{\theta}_j, M_j)}{\sum_{i=1}^{2^K} P\left(M_i\right)\left(NT\right)^{\frac{-k_i}{2}} f(data|\widehat{\theta}_i, M_i)} \tag{31}$$

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 (31) could also be obtained with a reasonable prior on the parameter space[33] that is known as Unit Information Prior (UIP). Moreover, Eicher et al. (2009b) conclude that this UIP combined with the uniform model prior (i.e. all models are equally probable a priori so that $P\left(M_j\right) = 1/2^K \forall j$) we consider in the paper outperforms any other possible combination of priors previously considered in the BMA literature

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

[32] 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 $V(\theta|data)_{uncond} = [V(\theta|data)_{cond} + E^2(\theta|data)_{cond}] \times PIP - E^2(\theta|data)_{uncond}$.

[33] 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.

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:

$$\text{PIP} = P\left(\theta_h \neq 0|data\right) = \sum\nolimits_{\theta_h \neq 0} P\left(M_j|data\right). \tag{32}$$

References

Acemoglu, D. (2009): *Introduction to Modern Economic Growth*, Princeton University Press.

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.

Akashi, K. and N. Kunitomo (2010): "The Limited Information Maximum Likelihood Approach to Dynamic Panel Structural Equations," *CIRJE Discussion Paper*, No. 708.

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. 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.

Barro, R. and X. Sala-i Martin (1992): "Convergence," *Journal of Political Economy*, 100, 223–251.

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., 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.

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

——— (2008): *Model Selection and Model Averaging*, Cambirdge University Press.

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.

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

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

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

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

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

Fernández, 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?" *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.

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

Hansen, B. and J. Racine (2010): "Jackknife Model Averaging," *Unpublished Manuscript*.

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," *The American Economic Review*, 82, 942–963.

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

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

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.

Masanjala, W. and C. Papageorgiou (2008): "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa using Bayesian Model Averaging," *Journal of Applied Econometrics*, 23, 671–682.

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

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.*

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

RODRIK, D., ed. (2003): *In Search of Prosperity: Analytic Narratives on Economic Growth*, Princeton University Press.

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.

SALA-I-MARTIN, X., G. DOPPELHOFER, AND R. MILLER (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *The American Economic Review*, 94, 813–835.

SARGAN, D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

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.

TSANGARIDES, C. (2004): "A Bayesian Approach to Model Uncertainty," *IMF Working Papers*, 04/68.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS[1]

1001 JAVIER ANDRÉS, ÓSCAR ARCE AND CARLOS THOMAS: Banking competition, collateral constraints and optimal monetary policy.
1002 CRISTINA BARCELÓ AND ERNESTO VILLANUEVA: The response of household wealth to the risk of losing the job: evidence from differences in firing costs.
1003 ALEXANDER KARAIVANOV, SONIA RUANO, JESÚS SAURINA AND ROBERT TOWNSEND: No bank, one bank, several banks: does it matter for investment?
1004 GABRIEL PEREZ-QUIROS AND HUGO RODRÍGUEZ MENDIZÁBAL: Asymmetric standing facilities: an unexploited monetary policy tool.
1005 GABRIEL JIMÉNEZ, JOSE A. LOPEZ AND JESÚS SAURINA: How does competition impact bank risk-taking?
1006 GIUSEPPE BERTOLA, AURELIJUS DABUSINSKAS, MARCO HOEBERICHTS, MARIO IZQUIERDO, CLAUDIA KWAPIL, JEREMI MONTORNÈS AND DANIEL RADOWSKI: Price, wage and employment response to shocks: evidence from the WDN Survey.
1007 JAVIER MENCÍA: Testing non-linear dependence in the Hedge Fund industry.
1008 ALFREDO MARTÍN-OLIVER: From proximity to distant banking: Spanish banks in the EMU.
1009 GALO NUÑO: Optimal research and development expenditure: a general equilibrium approach.
1010 LUIS J. ÁLVAREZ AND PABLO BURRIEL: Is a Calvo price setting model consistent with micro price data?
1011 JENS HAGENDORFF, IGNACIO HERNANDO, MARÍA J. NIETO AND LARRY D. WALL: What do premiums paid for bank M&As reflect? The case of the European Union.
1012 DAVID DE ANTONIO LIEDO: General equilibrium restrictions for dynamic factor models.
1013 JAMES COSTAIN, JUAN F. JIMENO AND CARLOS THOMAS: Employment fluctuations in a dual labor market.
1014 LUIS M. VICEIRA AND RICARDO GIMENO: The euro as a reserve currency for global investors.
1015 PALOMA LÓPEZ-GARCÍA AND JOSÉ MANUEL MONTERO: Understanding the Spanish business innovation gap: The role of spillovers and firms' absorptive capacity.
1016 AITOR LACUESTA AND SERGIO PUENTE: El efecto del ciclo económico en las entradas y salidas de inmigrantes en España.
1017 REBEKKA CHRISTOPOULOU, JUAN F. JIMENO AND ANA LAMO: Changes in the wage structure in EU countries.
1018 THOMAS BREUER, MARTIN JANDAČKA, JAVIER MENCÍA AND MARTIN SUMMER: A systematic approach to multi-period stress testing of portfolio credit risk.
1019 LUIS J. ÁLVAREZ AND PABLO BURRIEL: Micro-based estimates of heterogeneous pricing rules: The United States vs. the euro area.
1020 ALFREDO MARTÍN-OLIVER AND VICENTE SALAS-FUMÁS: I.T. investment and intangibles: Evidence from banks.
1021 LUISA LAMBERTINI, CATERINA MENDICINO AND MARIA TERESA PUNZI: Expectations-driven cycles in the housing market.
1022 JULIÁN MESSINA, PHILIP DU CAJU, CLÁUDIA FILIPA DUARTE, NIELS LYNGGÅRD HANSEN AND MARIO IZQUIERDO: The incidence of nominal and real wage rigidity: an individual-based sectoral approach.
1023 ALESSIO MORO: Development, growth and volatility.
1024 LUIS J. ÁLVAREZ AND ALBERTO CABRERO: Does housing really lead the business cycle?
1025 JUAN S. MORA-SANGUINETTI: Is judicial inefficiency increasing the house property market weight in Spain? Evidence at the local level.
1026 MAXIMO CAMACHO, GABRIEL PEREZ-QUIROS AND PILAR PONCELA: Green shoots in the Euro area. A real time measure.
1027 AITOR ERCE AND JAVIER DÍAZ-CASSOU: Creditor discrimination during sovereign debt restructurings.
1028 RAFAEL REPULLO, JESÚS SAURINA AND CARLOS TRUCHARTE: Mitigating the pro-cyclicality of Basel II.
1029 ISABEL ARGIMÓN AND JENIFER RUIZ: The effects of national discretions on banks.
1030 GABRIEL JIMÉNEZ, STEVEN ONGENA, JOSÉ-LUIS PEYDRÓ AND JESÚS SAURINA: Credit supply: identifying balance-sheet channels with loan applications and granted loans.
1031 ENRIQUE MORAL-BENITO: Determinants of economic growth: A Bayesian panel data approach.

---

1. Previously published Working Papers are listed in the Banco de España publications catalogue.

1032 GABE J. DE BONDT, TUOMAS A. PELTONEN AND DANIEL SANTABÁRBARA: Booms and busts in China's stock market: Estimates based on fundamentals.

1033 CARMEN MARTÍNEZ-CARRASCAL AND JULIAN VON LANDESBERGER: Explaining the demand for money by non-financial corporations in the euro area: A macro and a micro view.

1034 CARMEN MARTÍNEZ-CARRASCAL: Cash holdings, firm size and access to external finance. Evidence for the euro area.

1035 CÉSAR ALONSO-BORREGO: Firm behavior, market deregulation and productivity in Spain.

1036 OLYMPIA BOVER: Housing purchases and the dynamics of housing wealth.

1037 DAVID DE ANTONIO LIEDO AND ELENA FERNÁNDEZ MUÑOZ: Nowcasting Spanish GDP growth in real time: "One and a half months earlier".

1038 FRANCESCA VIANI: International financial flows, real exchange rates and cross-border insurance.

1039 FERNANDO BRONER, TATIANA DIDIER, AITOR ERCE AND SERGIO L. SCHMUKLER: Gross capital flows: dynamics and crises.

1101 GIACOMO MASIER AND ERNESTO VILLANUEVA: Consumption and initial mortgage conditions: evidence from survey data.

1102 PABLO HERNÁNDEZ DE COS AND ENRIQUE MORAL-BENITO: Endogenous fiscal consolidations.

1103 CÉSAR CALDERÓN, ENRIQUE MORAL-BENITO AND LUIS SERVÉN: Is infrastructure capital productive? A dynamic heterogeneous approach.

1104 MICHAEL DANQUAH, ENRIQUE MORAL-BENITO AND BAZOUMANA OUATTARA: TFP growth and its determinants: nonparametrics and model averaging.

1105 JUAN CARLOS BERGANZA AND CARMEN BROTO: Flexible inflation targets, forex interventions and exchange rate volatility in emerging countries.

1106 FRANCISCO DE CASTRO, JAVIER J. PÉREZ AND MARTA RODRÍGUEZ VIVES: Fiscal data revisions in Europe.

1107 ANGEL GAVILÁN, PABLO HERNÁNDEZ DE COS, JUAN F. JIMENO AND JUAN A. ROJAS: Fiscal policy, structural reforms and external imbalances: a quantitative evaluation for Spain.

1108 EVA ORTEGA, MARGARITA RUBIO AND CARLOS THOMAS: House purchase versus rental in Spain.

1109 ENRIQUE MORAL-BENITO: Dynamic panels with predetermined regressors: likelihood-based estimation and Bayesian averaging with an application to cross-country growth.

BANCO DE **ESPAÑA**
Eurosistema

Unidad de Publicaciones
Alcalá 522, 28027 Madrid
Telephone +34 91 338 6363. Fax +34 91 338 6488
E-mail: publicaciones@bde.es
www.bde.es