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

Enrique Moral-Benito†

November 2011

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 predetermined 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.

JEL Codes: C23, O40

Keywords: Dynamic Panel Estimation, Maximum Likelihood, Weak Instruments, Growth Regressions, Bayesian Model Averaging.

*All programs and data used in this paper together with replication instructions are available from my website http://www.moralbenito.com. In this website 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: Banco de España. 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-La Pietra-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.

1 Introduction

In this paper we consider a linear (dynamic) panel data model with general predetermined explanatory variables[1] 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. 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 over identifying 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. Other researchers such as Akashi and Kunitomo (2010) have also considered LIML estimators for such a panel model in the spirit of the continuously updated GMM estimators discussed in Hansen et al. (1996). 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

[1] Throughout the paper we also refer to a predetermined explanatory variable as a partially endogenous variable. Partial endogeneity can be understood as any intermediate configuration between strict exogeneity (i.e. the regressor is uncorrelated with past, present, and future shocks) and strict endogeneity (i.e. the regressor is correlated with all past, present, and future shocks).

effects have been discussed in the literature (e.g. Bhargava and Sargan (1983); Hsiao et al. (2002); 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.[2] 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 an alternative without resorting to auxiliary stationarity assumptions in the spirit of Arellano and Bover (1995) and Blundell and Bond (1998).

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 uncer-

[2]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.

tainty 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 under the assumption of partially endogenous regressors. The possibility to simultaneously address the problems of model uncertainty and endogeneity seems of paramount importance for empirical growth researchers.[3]

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. Lastly, Section 5 concludes. Additional results are gathered in a supplementary appendix to this paper.[4]

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).[5] 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

[3]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)

[4]Available at: http://www.moralbenito.com

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

general predetermined variables) allowing for feedback from lagged values of y to the current value for x. More precisely, assumption (2) allows x variables in period t to be 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 the likelihood framework we develop in this paper. For instance, we can also have non-zero correlations between the partially endogenous regressors —x_{it}— and contemporaneous shocks —v_{it}— (see below). We label the x variables as predetermined or partially endogenous[6] 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 the supplementary appendix.

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.[7]

Models like the one presented in equations (1)-(2) are typically estimated by first-differenced generalized method-of-moments. However, the conclusion from a sizable 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 resorting to auxiliary assumptions from a method-of-moments perspective (see for example Arellano and Bover (1995), Hansen et al. (1996),[8] 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

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

[7]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.

[8]Despite commonly-used in panel data models, note that the original continuously updated GMM estimators discussed in Hansen et al. (1996) were not explicitly proposed for panels.

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 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:[9]

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

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

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

[9]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.

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:

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

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:

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

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:

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

where R and X_t are the following matrices:

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

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

It is important to remark here that the maximizer of L is a consistent and asymptotically normal estimator regardless of non-normality. 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.1) 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 write:

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

Note that, again, in (6) we are implicitly assuming that $Cov(\eta_i, w_i) = 0$ in order to ensure identification of δ. The scalar γ_0 and the $k \times 1$ vector γ_0 represent parameters to be estimated, and ϵ_i can be interpreted as an individual-specific effect uncorrelated with the initial observations.

Moreover, by substituting (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:

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

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:

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

it is then possible to rewrite:

$$\left(\begin{array}{c|c} B_{11}^S & B_{12}^S \\ \hline 0 & I_{k-1} \end{array}\right) \left(\begin{array}{c} R_{i1}^S \\ \hline R_{i2}^S \end{array}\right) = \left(\begin{array}{c} \Pi_1 \\ \hline \Pi_2 \end{array}\right) z_i + \left(\begin{array}{c} U_{i1} \\ \hline U_{i2} \end{array}\right) \tag{9}$$

where:

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

$$\Pi_1 = \begin{pmatrix} \alpha+\gamma_0 & \beta'+\gamma'_1 & \delta' \\ \gamma_0 & \gamma'_1 & \delta' \\ \vdots & \vdots & \vdots \\ \gamma_0 & \gamma'_1 & \delta' \end{pmatrix}_{T\times(1+k+m)} \qquad \Pi_2 = \begin{pmatrix} \pi_{20} & \pi_{21} & \pi_2^w \\ \vdots & \vdots & \vdots \\ \pi_{T0} & \pi_{T1} & \pi_T^w \end{pmatrix}_{k(T-1)\times(1+k+m)}$$

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

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

where:

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

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

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

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

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

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

- Ω_{12}^S captures the feedback process. In particular, given the assumptions above 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:[10]

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

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

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

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.[11]

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.1 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 in the framework of empirical growth regressions.

3.1 Model and Estimators

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

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

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

[11]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: $(Y_{i1}', ..., Y_{iT}')'$ where $Y_{it} = (y_{it-1}, x_{it}^1, x_{it}^2)'$ and $(i = 1, ..., N)$. On the other hand, we assume that initial observations $Y_{i1} = (y_{i0}, x_{i1}^1, x_{i1}^2)'$ are observed. We further assume that the initial observations and the fixed effect are jointly normally distributed[12] 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.2.

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 Y 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:

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

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

[12]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 the supplementary appendix 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. We do not consider system-GMM competing estimators in the Monte Carlo exercise because they have not received much attention in the growth regressions literature given their required stationarity assumptions.

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 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

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.

this is an indication of weak instruments in the diff-GMM estimator. On the other hand, looking at the interquartile range (iqr), WG has always less dispersion than diff-GMM and sub-sys LIML as expected. However, the dispersion of sub-sytem LIML is very similar to that of diff-GMM and even smaller for the α parameter. This means that the higher probability of outliers in LIML estimators is not a big concern in this particular application. Finally, attending to MAE's, sub-sys LIML always performs clearly better than diff-GMM. MAE summarizes information on the performance of the estimator in terms of both bias and dispersion. Summing up, the conclusion from Panel A in Table 1 is that sub-system LIML clearly ourtperforms diff-GMM in the typical situation that an empirical growth researcher faces when using ten-year periods over the post-war sample 1960-2000.

In Panels B and C of Table 1, the results with $N = 500$ and $N = 1000$ are presented for illustrating the performance of the estimators in larger samples. In principle this is not a realistic situation in the cross-country growth context since there are not so many countries in the world. However, one could use regional data and have a sample size of a magnitude similar to 500 in the cross-section dimension. In any case, the purpose of this experiment is to investigate the relative performance of diff-GMM and sub-sys LIML in larger samples (larger in the cross-section dimension) since both estimators are consistent as $N \to \infty$ and T remains fixed. The performance of WG is not affected by increasing N since the WG bias comes from the small sample size in the time series dimension. Therefore, in terms of median bias, the WG results are practically the same in Panels A, B, and C. However, as expected, diff-GMM performance substantially improves as N increases in terms of median bias and dispersion. This improvement is not so substantial for sub-sys LIML since its performance is already reasonably satisfactory with $N = 100$ as shown in Panel A. However, looking at MAE's as a summary measure, sub-system LIML is still considerably better than diff-GMM in all cases. In any event, while sub-sys LIML biases become insignificant for moderate values of N, the diff-GMM biases are not negligible even with $N = 1000$. This would lead us to the conclusion that, with four time series observations, in order to consider the consistency results valid in this application, diff-GMM requires sample sizes larger than 1000 in the cross-section dimension, which seems clearly implausible in the growth context.

Three additional experiments based on $T = 8$ are presented in the three bottom panels of Table 1. I also consider these experiments because five-year periods are commonly considered in the panel growth literature, and, if we consider the post-war period 1960-2000, we would end up with eight time series observations. Panels D, E, and F present the results with $N = 100$, $N = 500$, and $N = 1000$ respectively. These results confirm the patterns previously described (i.e. sub-sys LIML clearly outperforms diff-GMM for all sample sizes in the cross-section dimension) but now, with $T = 8$, the biases and interquartile ranges for both diff-GMM and sub-sys LIML are always smaller for a given value of N. This means that the performance of both estimators clearly improves as the number of time series observations increases. As expected, this is also true in the case of WG.

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.2. 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, the supplementary appendix 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:[13]

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

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. See Appendix A.3 for a list of growth determinants considered in the paper and their sources. Moreover, the supplementary appendix provides a brief overview of previous findings in the literature regarding these growth determinants.

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 the supplementary appendix 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 the framework of single model approaches.

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

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. Therefore, researcher's uncertainty about the value of the parameter of interest in a growth regression exists at distinct two levels. The first one is the uncertainty associated with the parameter conditional on a given empirical growth model. This level of uncertainty is of course assessed in virtually every empirical study. What is not fully assessed is the uncertainty associated with the specification of the empirical growth model. It is typical for a given paper that the specification of the growth regression is taken as essentially known; while some variations of a baseline model are often reported, via different choices of control variables, standard empirical practice does not systematically account for the 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.[14] The fundamental principle of BMA is to treat both models and parameters as unobservable, and to estimate their distributions based on the observable data.[15] The basics of Bayesian model averaging are presented in the supplementary appendix.

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

[14] 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).

[15] 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.

with the same objective. However, both studies rely on the exogeneity assumption of the regressors and focus on cross-sectional data.[16] 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 strictly 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.[17]

In this section we combine BMA with the likelihood function previously introduced in order to simultaneously address partial endogeneity of the regressors and model uncertainty in the context of (panel) growth empirics. More specifically, we consider Unit Information Prior on the parameter space (i.e. Schwarz asymptotic approximation to Bayes Factors) and uniform priors on the model space (i.e. all models are equally probable). According to Eicher et al. (2009b) this combination of priors in the BMA framework identifies the largest possible set of growth determinants. The supplementary appendix provides more details on the BMA setting considered in this paper.

4.2 Empirical Results

Table 2 presents the results when combining the panel likelihood-based estimator presented in Section 2 with Bayesian model averaging techniques. In the context of empirical growth regressions, this combination represents an attempt to simultaneously consider model uncertainty and endogeneity of growth regressors.

Regarding the issue of convergence, the point estimate of the rate of convergence[18] 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.[19] This finding casts doubt on the conventional wisdom of conditional convergence as a

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

[17]Heuristically, these approaches consider pseudo likelihood functions replacing 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.

[18]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)).

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

Table 2: Empirical Results on Growth Determinants

	Posterior mean conditional on inclusion (1)	Posterior s.d. conditional on inclusion (2)	Fraction of models with $\|tstat\| > 2$ (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 described in the supplementary appendix. The sample covers the period 1960 to 2000 divided in 10-years sub periods. 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.

strong empirical regularity in the country-level data (e.g. Barro and Sala-i Martin (1992), Caselli et al. (1996)).[20] For illustrative purposes we plot in Figure 1 the BMA posterior distribution of the convergence coefficient which presents a substantial amount of probability mass on both sides of one.[21]

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

[21]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 dis-

Figure 1: Posterior Distribution of the Convergence Coefficient

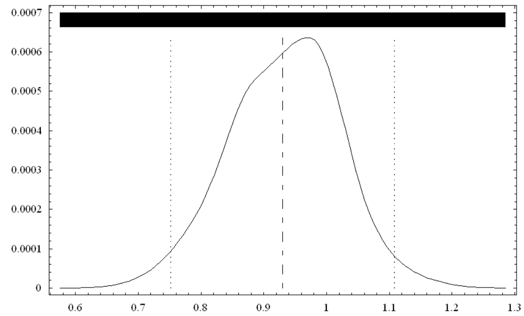

Figure 1 presents the marginal posterior distribution of the coefficient on the lagged dependent variable (i.e. the convergence coefficient). The graph consists of two parts: a gauge on top of the graph that indicates the Posterior Inclusion Probability (PIP) of the variable (which is 1 by definition 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.[22] 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

tribution 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.

[22]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.

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 in the paper (see Appendix A.3).

This result is in contrast to previous findings in the literature.[23] 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

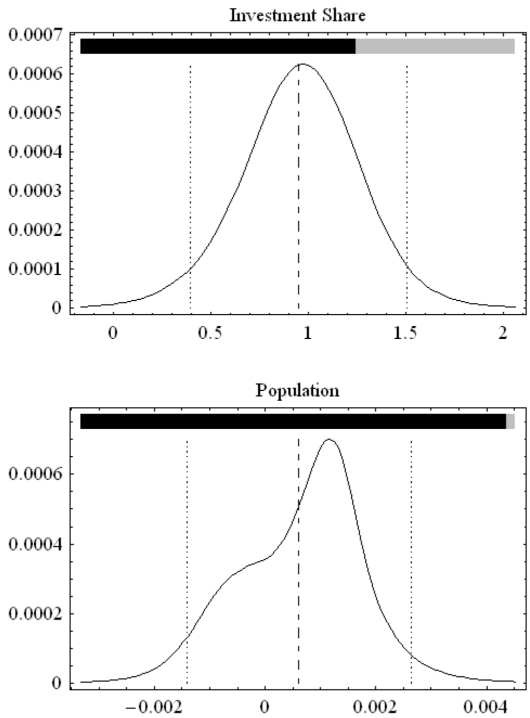

Figure 2 presents the marginal posterior distributions of the investment share and population coefficients. In particular, each graph consists of two parts: a gauge on top of the graphs that indicates the Posterior Inclusion Probability (PIP) of the variables and the normal mixture density for each coefficient's 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.

[23]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 the supplementary appendix). 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.[24] 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.

[24]We refer here to diff-GMM as standard GMM estimators since they are the most commonly-used in the growth regressions framework considered in the paper. Other GMM alternatives such as system-GMM (e.g. Blundell and Bond (1998); Arellano and Bover (1995)) are not explicitly discussed in this paper because their required stationarity assumptions are not very appealing in settings such as growth empirics.

A Appendix

A.1 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:

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

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:

$$\begin{aligned} L &\propto -\frac{N}{2} \ln \det \Omega_{11} + \frac{N}{2} \ln \det G_{22} - \frac{1}{2} tr(\Omega_{11}^{-1} U_1' U_1) - tr(F_{12} G_{22} U_2' U_1) \\ &- \frac{1}{2} tr(G_{22} U_2' U_2) - \frac{1}{2} tr(F_{12} G_{22} F_{21} U_1' U_1) \end{aligned} \quad (18)$$

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 (18) with respect to Π_2 and obtain its ML estimate:

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

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

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

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

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

Replacing in (18), we obtain L_2, the log-likelihood concentrated with respect to Π_2:

$$\begin{aligned} L_2 &\propto -\frac{N}{2}\ln\det\Omega_{11} + \frac{N}{2}\ln\det G_{22} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) \\ &- \frac{1}{2}tr\{(R_2 + U_1F_{12})'Q(R_2 + U_1F_{12})G_{22}\} \end{aligned} \tag{19}$$

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:

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

Therefore:

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

By differentiating the log-likelihood function, we obtain:

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

This implies that:

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

and so the final concentrated log-likelihood is:

$$L_3 \propto -\frac{N}{2}\ln\det\Omega_{11} - \frac{1}{2}tr(\Omega_{11}^{-1}U_1'U_1) - \frac{N}{2}\ln\det(\frac{1}{N}H) \tag{20}$$

A.2 Monte Carlo Details

For simulating the data in the Monte Carlo experiment, we first estimate a tri-variate VAR process for GDP[25] (y), investment ratio (x^1) and population growth (x^2). In particular, we consider the following VAR process:

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

where:

$$\begin{aligned}
Y_{it} &= (y_{it-1}, x^1_{it}, x^2_{it})' \\
\zeta_i &= (\zeta^y_i, \zeta^1_i, \zeta^2_i)' \\
\vartheta_{it} &= (\epsilon^y_{it}, \epsilon^1_{it}, \epsilon^2_{it})' \\
Var\left((Y'_{i1}, \zeta'_i)'\right) &= \Omega_{MC} \\
Var\left(\vartheta_{it}\right) &= \Sigma_{MC}
\end{aligned}$$

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

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

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

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

[25] 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.3 Data Appendix

Table A2: Variable Definitions and Sources

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

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

Table A3: List of Countries

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

References

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

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. and J. Lee (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.

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.

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.

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.

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.

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.

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.

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

Hsiao, C., H. Pesaran, and A. Tahmiscioglu (2002): "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

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.

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.

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.

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