INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des documents de travail
de la Direction des Etudes et Synthèses Économiques

G 2001 / 05

Testing the augmented Solow growth model: An empirical reassessment using panel data

Cédric AUDENIS *
Pierre BISCOURP *
Nathalie FOURCADE *
Olivier LOISEL **

AVRIL 2001

_____

*   Département des Etudes Economiques d'Ensemble - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

**   CEPREMAP - 142, rue du chevaleret - 75013 PARIS et London School of Economics

Testing the augmented Solow growth model:
An empirical reassessment using panel data

Abstract

We estimate a conditional convergence equation derived from an augmented Solow model where human capital is defined as skilled labour. We implement Generalized Method of Moments (GMM) estimators on a panel of countries. Estimation is carried out on the model in levels instrumented by the lagged first differences of the explanatory variables. We argue that using GMM estimators on the first differenced equation, as was previously attempted, is inappropriate as lagged levels of regressors provide weak instruments for current first differences. Using Asymptotic Least Squares, we check that the GMM estimates for structural parameters are consistent with the restrictions imposed by the model under the classical assumption of constant speed of convergence. We find that countries cover half the distance to their steady state in 16 years. Nonlinear estimations carried out under the assumption of endogenous speed of convergence do not validate the Solow model, whose theoretical predictions as to the dependence of the speed of convergence on structural parameters appear to be at best fragile.

Keywords: convergence, growth, generalized method of moments, dynamic panel data

Une ré-évaluation du modèle de Solow
sur données de panel

Résumé

Nous estimons sur un panel de 77 pays une équation de convergence conditionnelle, dérivée d'un modèle de Solow « augmenté » dans lequel le capital humain est défini comme le stock de travailleurs qualifiés. Afin de corriger des biais d'hétérogénéité inobservée et de simultanéité, nous utilisons la Méthode des Moments Généralisés. Nous montrons que la méthode traditionnelle consistant à estimer l'équation en différences premières en l'instrumentant par les niveaux passés des variables explicatives est ici inappropriée parce que les valeurs retardées des régresseurs sont des instruments faibles. Nous réalisons en conséquence l'estimation sur le modèle en niveau instrumenté par les différences premières retardées. A l'aide des Moindres Carrés Asymptotiques, nous vérifions que les coefficients estimés par cette méthode sont cohérents avec les restrictions théoriques imposées par le modèle sur les paramètres structurels, sous l'hypothèse standard de vitesse de convergence constante. Nous trouvons que les pays couvrent la moitié de la distance qui les sépare de leur état stationnaire en 16 ans. Les estimations non-linéaires menées sous l'hypothèse de vitesse de convergence endogène - pleinement cohérente avec le modèle théorique - ne permettent pas de trancher quant à la validité du modèle de Solow.

Mots-clés : convergence, croissance, méthode des moments généralisés

JEL classification : O41, O47

Introduction

The Solow model [1956] is generally viewed as a dinosaur in today's endogenous growth literature. Yet, it has won the status of standard growth model for international institutions' economists, because of its very simplicity which entails simple predictions. Considering data on macroeconomic growth have obvious limits, because the number of observations as well as the number of well-measured variables are scarce, this simplicity is synonymous of robustness: one may only hope to find empirically significant effects for first-order phenomena well captured by the basic Solow theoretical framework.

Several competing lines of research have developed during the last decade in order to assess the validity of the Solow model. One of them focuses on the issue of parameter heterogeneity across groups of countries and the importance of nonlinearities in the growth process (see for instance Quah [1997] and Temple [1999] for a review of the recent literature on the topic). Another approach takes a different view and postulates that the parameter homogeneity assumption, considered as an approximation to real situations, extends the possibilities of econometric estimation by allowing the use of panel data. In this paper, we follow the latter approach, arguing (see Temple [1999] for a detailed discussion) that both lines of research are worth following, as they point to complementary issues on the matter of explaining the process of growth. We do so because previous attempts seem to us not to have exploited all the methodological tools available in the field of the econometrics of panel data.

Following the line of research initiated by Mankiw, Romer and Weil [1992], we concentrate on the issue of convergence: the simple Solow model predicts that economies sharing the same saving rate and population growth rate should eventually converge to the same long run path given initial conditions. The issue of whether such a crude theoretical framework may be validated by the data has given rise to a huge empirical literature, spurred by the availability of the Summers and Heston [1988] dataset. Results so far appear to be inconclusive. Let us first summarize them briefly before presenting the methodological improvements we think may be made to this approach.

Mankiw, Romer and Weil [1992] proposed a test of the augmented[1] Solow model based on the estimation of a convergence equation. Their results were criticized, however, as doubts were shed on the use of Ordinary Least Squares (OLS) to estimate a dynamic equation on cross section data (Islam [1995]). Indeed, OLS yields unbiased estimates only as long as residuals are uncorrelated with explanatory variables. This condition may hold for the static steady state equation estimated by Mankiw Romer and Weil, as the endogeneity of regressors with respect to productivity shocks remains subject to discussion. However, estimating a static equation rests on the assumption that economies are likely to have reached their steady state within the estimation period. A conditional convergence framework is both more general and more convincing. OLS, however, is no more an appropriate method for estimating a convergence equation[2]: it seems unreasonable to extend the assumption of absence of correlation between the residual and exogenous regressors to the lagged endogenous variable, as the latter contains unobserved individual effects. Residuals derived from the Solow model in particular contain a country specific initial level of technology likely to be correlated with initial GDP levels.

In order to deal with this unobserved heterogeneity issue, Islam [1995] advocates the use of the Chamberlain[3] estimation method. However, the Π matrix approach he uses, as well as Knight, Loayza, and Villaneuva [1993], focuses on the issue of the endogeneity of regressors with respect to individual effects, and does not tackle the problem of the simultaneity of investment rates and productivity shocks.

---

[1] i.e. including human capital accumulation.
[2] i.e. containing an autoregressive term.
[3] For a presentation of the P matrix approach, see Chamberlain [1984].

Caselli, Esquivel and Lefort [1996] therefore implements Generalized Method of Moments (GMM) panel data estimators. They estimate a first differenced convergence equation instrumented by lagged levels of explanatory variables. This has indeed become the standard approach since Arellano and Bond's [1991] seminal paper. Their results prove imprecise, which seems to raise doubts as to the very possibility to validate a simple human capital augmented Solow model.

We use both Mankiw, Romer and Weil [1992] and Caselli, Esquivel and Lefort [1996] as starting points for our empirical analysis of the issue of convergence. The former provides the general theoretical framework, used by all followers in this literature, which we shall discuss and adapt. The latter provides what seems to us to be the most rigorous econometric treatment available in the existing literature. This paper must be seen as an attempt to provide a more rigorous methodology on both aspects in order to assess whether the inconclusive results obtained so far (in this respect, Caselli, Esquivel and Lefort [1996] will be our benchmark) are due to the inadequacy of the augmented Solow model, to methodological problems within the line of research followed (that of panel data estimation under the assumption of homogenous technological parameters), or to the very approach. It must therefore be seen as a twofold methodological contribution.

Our estimation strategy first differs from previous attempts. We implement GMM panel data estimators on a model in levels instrumented by past first differences of the explanatory variables. This approach is valid, as Arellano and Bover [1995] have shown, under the additional identifying assumption of constant correlation between explanatory variables and the individual effect. We focus on this estimation strategy because given the high degree of autocorrelation of explanatory variables in a panel framework, lagged differences of regressors provide better instruments for current levels than lagged levels do for current first differences.

We also propose an alternative way of taking into account human capital accumulation, traditionally modelled by analogy with physical capital: since Mankiw, Romer and Weil [1992] a « saving rate » in human capital is applied to national income, the measure of this saving rate being provided by rates of schooling weighted by shares of population of schooling age. This approach is questionable since the very definition of human capital remains inexplicit. We define human capital as skilled labour. We compute a rate of investment in human capital derived from the share of total population who has attained a secondary level. In order to guarantee the theoretical consistency of the model, we apply this rate of investment to the country's population as opposed to its GDP.

Our estimates of structural parameters under the traditional assumption of constant speed of convergence yield plausible values: the elasticity of production to physical capital roughly equals 30%. We find a half-life of convergence equal to 16 years[4]. Our model is overidentified in that we estimate more coefficients than we have structural parameters. Using Asymptotic Least Squares, we prove that the constraints implied by the model are not rejected. It turns out that these results do not require the use of all past first differences as instruments.

Testing the Solow model under the assumption of constant speed of convergence is handy because it relies on linear econometrics methods. However, it is not theoretically consistent as deriving a convergence equation from an augmented Solow model yields an endogenous expression for the speed of convergence. We test a more general model where we allow for the presence of both a constant and an endogenous term in the speed of convergence. The endogenous component appears to be significant yet weakly consistent with the data, while the constant one appears to be significant, in contradiction with the model. We conclude that our econometric methodology allows to obtain sharp improvements in the accuracy of the estimated parameters, but that these results do not allow to conclude to the full consistency of the theoretical model with the data. They do not

---

[4] Thus, we find convergence to be faster than was estimated by Mankiw, Romer and Weil, yet slower than in Caselli, Esquivel and Lefort's study.

validate the use of the assumption of constant speed of convergence either, as the endogenous component of the latter is significant.

The remainder of the article is organized as follows. Section 1 briefly reviews previous attempts, points to their limits and presents our estimation strategy. The Solow convergence equation based on our modelling of human capital accumulation is derived in section 2. We then discuss our results under both assumptions of constant and endogenous speed of convergence.

Failures to validate the Solow model: wrong model or poor methodology?

Mankiw, Romer and Weil's [1992] validation of the Solow model rests on strong econometric assumptions

Mankiw, Romer and Weil (henceforth MRW) proposed in 1992 a simple test of the Solow model. The authors consider the basic textbook Solow model as well as an « augmented » version which includes human capital as an input of the production function. Their framework allows for conditional convergence, where saving and population growth (and consequently the long run path) are exogenous and idiosyncratic.

They first estimate a static equation under the assumption that the economy reaches its steady state within the period 1960-1985. Then, turning to conditional convergence, they estimate an autoregressive model. The first estimation by OLS is only valid if the regressors are uncorrelated with the residuals. Recall however that using OLS to estimate an autoregressive model on individual data is poor methodology leading to a downward bias on the estimate of the speed of convergence (Islam [1995]).

The model

The production function is a Cobb-Douglas with constant returns to scale:

$$Y_t = K_t{}^{\alpha} H_t{}^{\beta} \left(A_t L_t\right)^{1-\alpha-\beta} \qquad 0 < \alpha \qquad 0 \le \beta \qquad \alpha + \beta < 1$$

Y, *K*, *H* and *L* are respectively the production level, the stock of physical capital, the stock of human capital and labour[5]. Labour *L* and level of technology *A* grow exogenously at rates *n* and *g*.

Income per capita converges towards a long run path defined by:

$$\ln\left(\frac{Y_t}{L_t}\right) = \ln A_0 + gt + \frac{\alpha}{1-\alpha-\beta}\ln s_k + \frac{\beta}{1-\alpha-\beta}\ln s_h - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+\delta+g) \qquad (1)$$

s_k and s_h are respectively the saving rates in physical and human capital[6]. Both physical and human capital are assumed to depreciate at the same rate[7] δ.

The linearized convergence path towards the long run trajectory can be written as:

$$\ln\left(\frac{Y_t}{L_t}\right) = (1-e^{-\lambda t})\ln A_0 + gt + (1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\ln s_k + (1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\ln s_h$$

$$-(1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n+\delta+g) + e^{-\lambda t}\ln\frac{Y_0}{L_0} \qquad (2)$$

---

[5] Since the Solow model does not distinguish between total and working-age population, MRW use World Bank data on the population aged between 15 and 64 for their measure of *L*.

[6] Both are exogenous. s_h is obtained by multiplying the schooling rate of the population aged between 12 and 17 (taken from the UNESCO yearbook) by the share of the population aged between 15 and 19 in the working-age population.

[7] A value of 0.05 is henceforth assumed for $\delta + g$.

where $\lambda = (1-\alpha-\beta)(n+\delta+g)$ is the speed of convergence[8].

Estimation by Ordinary Least Squares

Both static steady state equation (1) and convergence equation (2) are estimated using the Summers and Heston dataset. Estimating equation (1) rests on the assumption that all countries have reached their steady state in 1985, starting from 1960. Steady states are idiosyncratic and defined by the corresponding values of n, s_k and s_h, computed as the average of these variables over the period 1960-1985[9].

Equation (2) by contrast allows for only partial convergence. Note however that estimation is carried out under the assumption that the speed of convergence λ may be estimated in the same way as technological parameters α and β. This is indeed the traditional approach to statistical convergence tests. However, one should be clear as to the goals assigned to the estimation of equation (2): if such estimations aim at providing a test of the Solow model, then all theoretical relationships derived from the model should be taken into account, in particular the endogenous expression of the speed of convergence. We turn to this criticism in more details in section 3, and focus in this section on the discussion of the estimation method used by MRW.

The equation residual reflects initial levels of technology: MRW split up the term $(1-e^{-\lambda t})\ln A_0$ appearing in equation (2) into a constant a and a country specific effect ε_A: $(1-e^{-\lambda t})\ln A_0 = a + \varepsilon_A$.

Specifying the model means making identifying assumptions on the properties of this residual. ε_A is assumed by MRW to be not only independent and identically distributed, but also orthogonal to all explanatory variables. This strong assumption allows them to use OLS.

Indeed, the static model described by equation (1) yields parameter estimates consistent with plausible values ($\alpha = \beta = 0.3$). The convergence equation (2), however, leads to far less satisfactory results: the estimates for α, β and λ respectively take the values 0.5, 0.2 and 0.014, the latter implying a fairly long half-life of convergence equal to 52 years[10].

Islam [1995] argues that this very low speed of convergence was to be expected, since the non-observed individual disturbance term ε_A is likely to be correlated with the initial state of the economy: as a result, OLS estimators are biased, even under the assumption of exogenous steady state explanatory variables, since the lagged endogenous variable is affected by the individual shock.

The panel data approach, once promising, yet apparently inconclusive

Islam [1995] advocates to this end the use of panel data. However, the Chamberlain method he uses does not tackle the issue of simultaneity. Caselli, Esquivel and Lefort [1996] (henceforth CEL) implements Generalized Method of Moments estimators on panel data in order to deal with both sources of bias. Their results prove inconclusive, which casts doubts on the validity of the whole approach.

---

[8] For a proof, see Mankiw, Romer and Weil [1992] or the appendix of this paper in a more general case.

[9] t is therefore a constant equal to 25, since the estimation period is 1960-1985.

[10] i.e. it takes 52 years for half the gap to the steady state to vanish.

Caselli, Esquivel and Lefort's approach [1996]:

The estimated equation

CEL extend the theoretical approach of MRW to a panel framework. They divide the 1960-1990 period into six five-year sub-periods. This allows to use panel data. The number of cross section units is large relative to the number of periods, which is the usual condition for panel data estimators to perform well.

Equations (1) and (2) become[11]:

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = a + \frac{\alpha}{1-\alpha-\beta}\ln s_{k_{it}} + \frac{\beta}{1-\alpha-\beta}\ln s_{h_{it}} - \frac{\alpha+\beta}{1-\alpha-\beta}\ln(n_{it}+\delta+g) + \varepsilon_{it} \qquad (1')$$

$$\ln\left(\frac{Y_{it}}{L_{it}}\right) = a + e^{-5\lambda}\ln\left(\frac{Y_{it-1}}{L_{it-1}}\right) + \left(1-e^{-5\lambda}\right)\frac{\alpha}{1-\alpha-\beta}\ln s_{it}^{k} + \left(1-e^{-5\lambda}\right)\frac{\beta}{1-\alpha-\beta}\ln s_{it}^{h}$$
$$-\left(1-e^{-5\lambda}\right)\frac{\alpha+\beta}{1-\alpha-\beta}\ln(n_{it}+\delta+g) + \varepsilon_{it} \qquad (2')$$

Again, CEL ignore the theoretical definition of the speed of convergence and stick to the traditional assumption that λ is constant across time and countries.

CEL transpose the MRW framework to each sub-period, so that one steady state is then available for each sub-period. This (now temporary) steady state undergoes changes from one sub-period to the next, caused by stochastic shocks assumed to be exogenous (i.e. independent of the past states of the economy). Because economies are unlikely to achieve full convergence towards their local steady state within each sub-period, equation (1') becomes irrelevant and only equation (2') remains to be estimated.

$\ln(n+g+\delta)$, $\ln(s_k)$ and $\ln(s_h)$ at time t are computed as the average of these quantities over the sub-period beginning in t. Note that contrary to MRW who base their definition of steady state on the average of explanatory variables over the convergence period, CEL use their current value: this implies that convergence takes place between $t-1$ and t towards a steady state defined by the vector of explanatory variables at date t.

Identifying restrictions

Residuals are decomposed into the following way:

$$\varepsilon_{it} = u_i + \eta_t + v_{it}$$

where u_i and η_t correspond respectively to the individual and time effects[12].

For an autoregressive equation, the OLS estimator is biased because the lagged endogenous variable included in the set of regressors is correlated with u_i (unobserved heterogeneity bias).

---

[11] Due to the panel structure of the data, equation residuals now include time effects as well as individual effects.

[12] CEL eliminate the time effect by centering the data by their mean over the sub-period. Note that unlike the first two components of the residual, v_{it} can not be derived from the model without further assumptions. Assume for instance that the level of technology $\ln A$ is randomly distributed across countries and through time, its expectation at time t being given by $\ln A_0 + gt$, we have : $\ln A_{it} = \ln A_{i0} + gt + v_{it}$.

Another source of bias stems from the likely correlation between v_{it} and the explanatory variables (endogeneity bias).

In order to deal with both sources of bias, CEL use the Generalized Method of Moments (GMM) under the assumption of weak exogeneity[13]. The economic intuition behind the latter is that current explanatory and endogenous variables are not correlated with future shocks, but may be correlated with past or current shocks. Indeed, one does not expect an unforecasted climatic shock at date $t+1$ to influence GDP, the population growth rate or the saving rates at date t. However, it is far less clear that this hypothesis of independence should also apply to current and past shocks.

Explanatory variables taken previous to the shock are therefore instruments. Current and future values however are not included in the set of instruments, as they are likely to be correlated with current shocks.

In practice CEL estimate the equation in first differences using the explanatory variables in levels lagged twice or more as instruments. This has become the standard approach since Arellano and Bond's seminal paper (1991).

Results: a relative failure

The estimation of the model leads to a negative β and to an implausible value for α (0.49), the speed of convergence λ being estimated at around 10% which implies a half life of convergence equal to 7 years. If CEL's methodology is to be taken seriously, these results undoubtedly argue against the Solow model.

An alternative econometric strategy

The relevance of CEL's methodology may however be questioned for two reasons:

- Their modelling of human capital, inherited from MRW, relies on a perfect symmetry between the processes of human and physical capital accumulation. This simplifies calculations but does not account for the singularity of human capital. This calls for an alternative modelling (see section 2).

- Estimating equation (2) written in first differences using instruments in levels may not be appropriate considering the type of variables we are dealing with.

We illustrate here the latter point.

Estimating the convergence equation in levels, instrumenting by lagged first differences performs better under serial correlation of some regressors.

Recall that two properties are required for variable Z to instrument explanatory variable X properly:

- Z must not be correlated with the residual of the model

- Z must be correlated with X, more precisely the coefficient of the regression of X on Z must be significantly different from zero. Whenever several instrumental variables are available, the most efficient (i.e. which leads to a minimal variance for the instrumental variables (IV) estimator) is the one best correlated with the explanatory variable.

---

[13] Namely $E(x_{it}u_i) \neq 0, E(x_{it}v_{is}) = 0 \quad \forall (t,s) / s > t$. Recall that the OLS estimator is biased even on the model written in first differences as long as the estimated equation is autoregressive.

Following Blundell & Bond [1998], we argue that lagged levels of explanatory variables are weak instruments for their first differences, whereas lagged first differences are appropriate instruments for levels, when the *X* variable exhibits strong enough serial correlation. We show this in the simple case where *X* follows an autoregressive process of order one: using lagged first differences of *X* to instrument current levels leads to a smaller variance of the IV estimator than using lagged levels of *X* to instrument current first differences.

Start from the following fixed effect model:

$$Y_{it} = \beta X_{it} + U_i + \varepsilon_{it}$$

with $V(\varepsilon_{it}) = \Sigma^2 I$ and $E(X_{it} U_i) \neq 0, E(X_{it}\varepsilon_{it}) \neq 0$.

Assume *X* follows an autoregressive process of order one:

$$X_{it} = \rho X_{it-1} + u_i + \omega_{it}$$

where $|\rho| < 1$, ω is a white noise (variance σ^2). Under these assumptions and the following stationarity condition

$$\text{cov}(X_{it}, u_i) = \text{cov}(X_{it'}, u_i) \quad \forall(t, t')$$

it is straightforward to obtain:

$$V(X) = \frac{\sigma^2}{1-\rho^2} + \frac{V(u_i)}{(1-\rho)^2}$$

$$V(\Delta X) = \frac{2\sigma^2}{1+\rho}$$

We want to compare the performances of two competing instrumental variable estimators of β:

- The estimator based on the model written in first differences, where the explanatory variable lagged twice provides an instrument for the current first difference

- The estimator based on the model in levels, where the first difference of the explanatory variable lagged once instruments the current level of *X*.

More precisely, we want to evaluate the ratio of the variances of both estimators as a function of the inertia of the process that generates *X*.

The asymptotic variances of the corresponding IV estimators have the following expressions:

$$V(\hat{\beta}_{IV}^{lev}) = V(\eta_{it}) \frac{V(\Delta X_{it-1})}{N \text{cov}^2(\Delta X_{it-1}, X_{it})}$$

$$V(\hat{\beta}_{IV}^{dif}) = V(\Delta \eta_{it}) \frac{V(X_{it-2})}{N \text{cov}^2(X_{it-2}, \Delta X_{it})}$$

where $\eta_{it} = U_i + \varepsilon_{it}$

Let Q denote the ratio of both quantities:

$$Q = \frac{V(\eta_{it})}{V(\Delta\eta_{it})} \frac{\text{cov}^2(X_{it-2}, \Delta X_{it})}{\text{cov}^2(\Delta X_{it-1}, X_{it})} \frac{V(\Delta X_{it-1})}{V(X_{it-2})}$$

The ratio of the squared covariances involved in the previous expression is easily shown to be one.

Q has therefore the following expression:

$$Q = \frac{\Sigma^2 + V(U_i)}{2\Sigma^2} \frac{2\sigma^2}{\dfrac{1+\rho}{1-\rho}\left[\dfrac{V(u_i)}{1-\rho} + \dfrac{\sigma^2}{1+\rho}\right]}$$

Let us characterize Q in terms of ρ and a parameter α equal to the ratio of the variance of the individual effect to the variance of the perturbation within the process generating X:

$$\alpha = \frac{V(u_i)}{\sigma^2}$$

Assume further for simplicity that this ratio is identical for the individual effect and the perturbation involved in the estimated equation. One then obtains a simple expression for Q:

$$Q = \frac{(1+\alpha)(1-\rho)^2}{1-\rho+\alpha(1+\rho)}$$

The distribution of Q given ρ and α is shown in Figure 1. When X exhibits no serial correlation (i.e. $\rho = 0$), Q is equal to one, in other words both estimators perform equally well or badly. However, for larger values of ρ, which are to be expected with macroeconomic data, the ratio of variances decreases sharply, all the more so when α is large, which is also to be expected.

**Figure 1 - Ratio of the variances of both instrumental variables estimators
in terms of ρ and α**
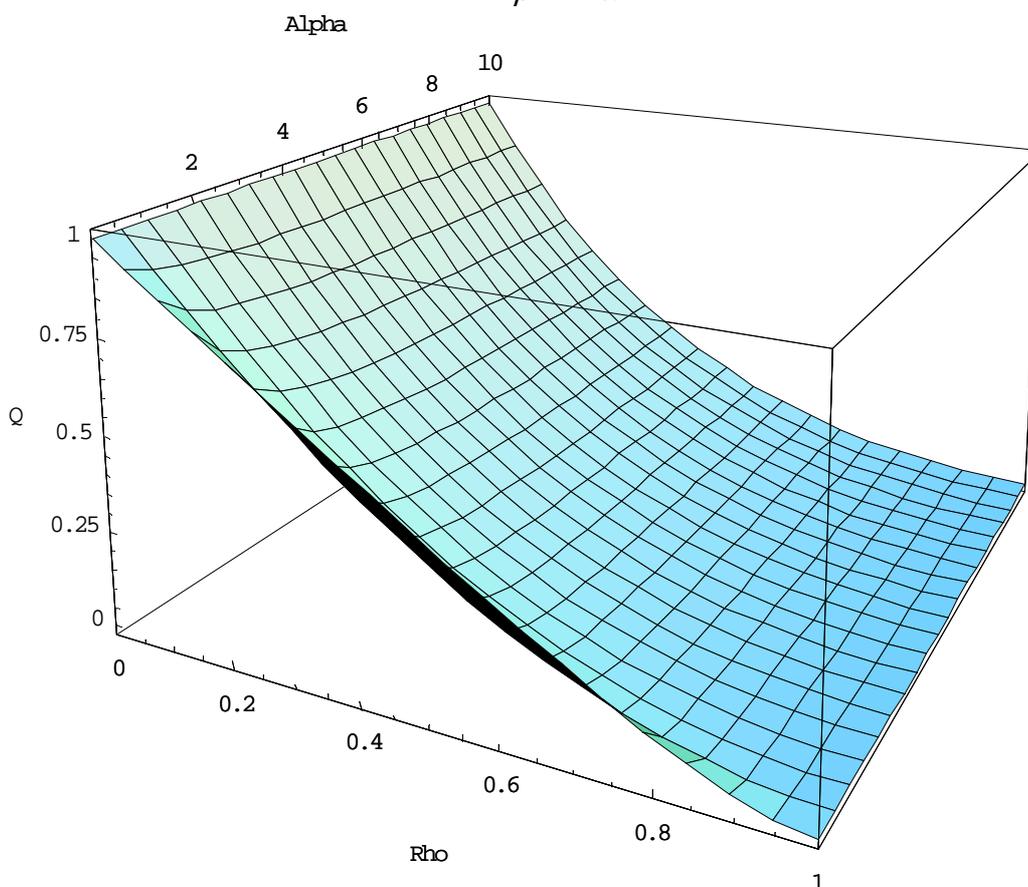
The previous considerations help understand why estimations of the first-differenced model perform badly when explanatory variables exhibit strong serial correlation. They also point to a remedy to this problem, which consists in estimating equation (2) in levels using first differenced explanatory variables lagged once or more as instruments.

However, equation (2') written in levels contains the individual effect u_i, which implies that orthogonality conditions derived from this specification are only valid under the additional condition $E(\Delta x_{it} u_i) = 0 \; \forall t$. Arellano and Bover [1995] first pointed that a sufficient condition for this equality to hold is the following stationarity assumption:

$$E(x_{it} u_i) = \delta$$

i.e. the correlation between the individual effect and the explanatory variables is not time dependent. This condition is intermediate between the absence of fixed effect ($\delta = 0$) and the less demanding assumption made by Arellano and Bond [1991] and used by CEL ($\delta = f(t)$).

Note that this assumption of constant correlation between explanatory variables such as saving rates, and the individual effects, may be invalid if the countries starting from a low technological level turn out to be the ones whose saving rates grow fastest (i.e. absolute convergence takes place). However, most of this absolute convergence has been taking place between groups of countries (such as the NIC and the OECD, the former tending to catch up with the latter). Hence, we deal with this criticism by centering all variables by date and group of countries, which is equivalent to adding a complete set of crossed

group and time dummies[14]. As in our framework country specific effects are deviations from group averages, the stationarity assumption may still be considered to hold.

The conditions under which the lagged first differences of the endogenous variable may be used as instruments can be derived by iterating equation (2') in levels. Blundell and Bond [1998] show that this is the case if the generating process described by equation (2') is convergent, goes back to infinity, and if the stationarity and weak exogeneity assumptions hold, as well as the absence of autocorrelation for v.

We proceed as follows:

- We start by checking for the absence of autocorrelation on the residual v of the model in levels. Crossing time and group effects reduces the risks of serial correlation on the residual v (which becomes a deviation from the group's average shock). There is still a need to test for autocorrelation. As the presence of an individual effect in the overall level residual inevitably makes it serially correlated, we estimate the first differenced equation using the endogenous and the explanatory variables in levels lagged twice or more as instruments. We check for autocorrelation of $v_{it} - v_{it-1}$ at order one and two. One expects autocorrelation at order one but not at order two if v is not serially correlated.

- The equation in levels is estimated using the endogenous and the explanatory variables in first differences lagged once or more as instruments. A Sargan test is used to check for the consistency of the set of orthogonality conditions (recall that a Sargan statistics follows a chi-squared with a number of degrees of freedom equal to the excess of orthogonality conditions over the number of parameters of interest).

This discussion finally leads to two theoretical specifications[15]:

Specification	Equation in level at time t	Equation in first differences at time t
Available instruments	$\Delta x_{i1},..., \Delta x_{i,t-1}$ $\Delta y_{i1},..., \Delta y_{i,t-1}$	$x_{i1},...,x_{i,t-2}$ $y_{i1},...,y_{i,t-2}$

The validity of instruments added to the set of instruments of any specification may be tested by a difference Sargan test: if the additional moment conditions are valid, then the difference between the Sargan statistics follows a χ^2 with a number of degrees of freedom equal to the number of additional moment conditions[16].

In the next section we present the model we estimate through these specifications.

---

[14] Time effects η are therefore eliminated in the process. We also tackle the issue of « group effects » pointing to strong observable heterogeneity, for instance between the OECD countries and the Newly Industrialized Countries versus the rest of the world.

[15] We have implemented a third specification called System Estimator, first suggested by Arellano and Bover [1995]. It consists in combining both specifications by adding a set of orthogonality conditions derived from the equations in levels, to the ones obtained from the model written in first differences. Results are not displayed here because the relevance of the method can be questioned in our case. Retaining all possible instruments may not be optimal when dealing with a small sample (see Appendix). Moreover, since the relevance of the orthogonality conditions corresponding to the first-differenced model is questionable, adding these conditions to the ones generated by the model in levels is of little interest. Indeed, results with this specification are more satisfactory than with System Estimator.

[16] This is valid provided the additional instruments are orthogonal to the initial set of instruments.

A Solow model with human capital defined as skilled labour

Defining and measuring human capital

Estimating an augmented Solow model has always been a tricky issue since it implies that one has in hand both a clear definition of the rate of investment in human capital, and a measure consistent with the theoretical concept. Previous studies have typically focused on the rate of secondary schooling, weighted by the share of population of schooling age, as this information is readily available from UNESCO for a large sample of countries (see MRW). This approach makes sense since it captures the intuition that the relevant issue when it comes to human capital accumulation is that of improving the available number of skilled workers in the economy.

However, the issue we want to deal with is the consistency of this measure of s_h with the theoretical framework it is embedded in. MRW for instance model human capital accumulation by analogy with physical capital: a share s_h is assumed to be devoted to human capital accumulation in the same way as a proportion s_k of national income is directed towards physical capital accumulation. This analogy is very handy when it comes to solving the model, since it allows one to make use of the symmetry of accumulation equations to derive an extended expression for the convergence path to the steady state, from the simple Solow convergence equation. However, applying a concept of human capital investment based on a rate of schooling to a financial income, may be viewed as dubious. Indeed, one would rather expect the share of education expenditure in GDP as a measure of s_h[17].

If one is concerned about modelling human capital accumulation in a way that makes it analogous to physical capital, a better choice for the corresponding concept of « income » would be the amount of total population for whom the decision is to be made at the beginning of every period, whether they should be schooled (i.e. « saved » to increase productivity next period), or not (i.e. « consumed » immediately as unskilled workers).

Contrary to CEL we define the population as the total population (and not only people of working age), since we expect this variable to be better measured. Recall that the Solow model does not distinguish between the two.

We define the stock of human capital by the number of working-age people who have attained the secondary level (this is a proxy for skilled labour). The rate of investment in human capital is then given by the accumulation equation:

$$H_{t+1} - H_t = s_t^h N_t - \delta H_t$$

The variation between t and t+1 in the stock of skilled people is due to the difference between the inflow of students graduating from high school, and an exit rate δ assumed to be constant. Note that by definition, s_h is applied to the stock of population at a given date. In order to interpret it as an investment rate, one would expect s_h to apply to a flow of population eligible for schooling, for instance the number of people aged between 15 and 20 years. Such data are to our knowledge not available. We consequently assume that there is some unobserved variable equal to the share of the 15-20 years in total population, which may vary from one country to another, and for a given country from one date to the next. As this variable is not measured, but is applied in a multiplicative way to the stock of total population, its logarithm ends up in the residual of our equation, either as part of the individual effect or as part of the v_{it} residual.

---

[17] The latter is a poor proxy for human capital investment, since it is polluted by unobserved heterogeneous efficiency across schooling systems. We therefore focus on their measured output.

We calculate our rate of human capital investment s_h as follows. We obtain the share R_t of total population N_t that has attained the secondary level at time t from the Barro and Lee dataset. Hence :

$$s_t^h = R_{t+1} \frac{N_{t+1}}{N_t} - (1-\delta)R_t \quad {}^{18}$$

The model

We specify the aggregate production function as:

$$Y(t) = A(t)K(t)^\alpha H(t)^\beta L(t)^{1-\alpha-\beta}$$

where $H(t)$ and $L(t)$ stand for skilled and unskilled labour and $A(t)$ is technical progress[19].

By assumption[20] :

$$N = H + L$$

We derive the human capital accumulation equation from the definition of s_h :

$$\dot{H} = s_h N - \delta_H H$$

Note that there is no explicit feedback of income on capital accumulation. Recall that the basic Solow model takes the saving rates as given exogenously. We extend this assumption to human capital investment which undergoes exogenous shocks every five years, causing the steady state of the economy to change.

Accumulation equations per head of population are then written:

$$\dot{k} = s_k y - (n + \delta_K)k$$

$$\dot{h} = s_h - (n + \delta_H)h$$

$$\dot{l} = -\dot{h}$$

As the ratio of output to the stock of physical capital is fairly constant over time, the rates of growth of per head output and per head capital must be identical at the steady state. We also assume that the per head skilled workforce is constant. Note that it follows from our definition of human capital that all variables do not grow at the same rates at the steady state. For this reason, we shall term it « limit trajectory ».

---

[18] Contrary to the other explanatory variables, which are calculated as the means over five years of their annual logarithm, the rate of investment in human capital is only known every five years from the Barro&Lee data set. This value is assumed to remain constant over the five year period following date t, and its annual equivalent is obtained from this expression divided by five. Moreover, the way s_h is computed implies that its value is missing for the most recent period.

[19] Recall that in the special case of a Cobb-Douglas function, all forms of technical progress are equivalent. Contrary to MRW and CEL, we start from a macroeconomic production function incorporating a Solow neutral technical change. The reason for this is that we define human capital as skilled labour. Since total population N is now divided into two components, it is no longer natural to model a Harrod neutral technical change and define all variables per efficient unit of labour. We therefore stick to per head values.

[20] This is obviously an approximation. If we assume, however, a share of aggregate labour in population constant across time, yet allowed to differ between countries, we only end up adding another individual effect to the residual of our equations. We deal with this effect as before. Alternatively, this approximation can be avoided by using the working-age population as a measure of N. Comfortingly, results are unchanged (results of the estimations using working-age population are not displayed here).

The method we use to derive a convergence equation to the limit trajectory defined above i.e. the path followed by per head output to its target when both values are initially different, consists in approximating accumulation equations around a steady state (*) obtained by transforming per head variables into another set of variables which have the property that they are constant when the limit trajectory is reached (as a result, skilled and unskilled labour are invariant by this transformation). We then plug these approximations into the log-differentiated production function and factorize this expression into a first order differential equation whose solution is the convergence path[21].

The solution (see appendix for a proof) is given by:

$$\ln y = e^{-\lambda t}\ln y_0 + \left(1 - e^{-\lambda t}\right)\ln y^*$$

$$\lambda = \left(1 - \alpha\right)\left(n + \delta\right)^*$$

This solution may be interpreted as follows. If at a given date taken as the origin of time, output per head happens to differ from the steady state implied by the (exogenous) values of saving rates and the growth rate of population for a given country, the Solow model predicts that convergence towards this steady state will take place at the speed λ. More precisely, output per head at any given date t along this convergence process is a weighted average of the initial value y_0 and of the steady state value (the « target » of convergence) y^*, with the weight on the initial value going to zero when t goes to infinity, the weight on the steady state value conversely going to one, all the more rapidly so if λ is large.

Mankiw, Romer and Weil [1992] estimated this equation for a cross section of countries between 1960 and 1985: in their case y_0 denotes output per head in 1960 for a given country, y denotes output per head measured in 1985, and y^* is assumed to be the steady state implied by the Solow model, computed by using averages of saving rates and growth rates of population taken over the whole 1960-1985 period. Indeed, under this assumption, the length of time seems to be sufficient to allow full convergence to steady states. One should bear in mind however that the definition of steady state they use is fully arbitrary. Our panel data framework is a generalization of that used by MRW. Convergence, likely to be partial, takes place over five or ten years: y_0 denotes output per head at the beginning of any of these periods of convergence, y denotes output per head measured at the end of the period, and y^* is assumed to be the steady state implied by the Solow model, computed by using averages of saving rates and growth rates of population taken over the period. Steady states are therefore fully exogenous, as in the MRW framework, but undergo stochastic shocks every five or ten years. The target towards which (partial) convergence takes place during a given period, therefore moves from one period to the next.

Expanding the previous expression using the definition of the steady state, we obtain the following equation, where the parameter τ is introduced to determine the length of the lag on the autoregressive term.

$$\ln y_t = (1 - e^{-5\lambda\tau})\frac{\alpha}{1-\alpha}\ln\left(s_t^k\right)^* + (1 - e^{-5\lambda\tau})\frac{\beta}{1-\alpha}\ln\left(s_t^h\right)^* - (1 - e^{-5\lambda\tau})\frac{1}{1-\alpha}\ln(n_t + \delta)^*$$

$$+ (1 - e^{-5\lambda\tau})\frac{1-\alpha-\beta}{1-\alpha}\ln(n_t + \delta - s_t^h)^* + e^{-5\lambda\tau}\ln y_{t-\tau} + \varepsilon_t \tag{2''}$$

---

[21] Note that this dynamics is one-dimensional. The literature on convergence within the Solow framework has to make assumptions on the values taken by the rates of depreciation of physical and human capital in order to obtain a factorisation of this dynamics. MRW for instance assumed the equality of both rates. Given our more general model, we must assume that the rate of depreciation of human capital is equal to that of physical capital plus the rate of growth of output on the limit trajectory.

Equation (2'') describes the path followed by a country while converging towards its steady state. Theory does not, however, imply any particular value for this convergence period, as identical values should be obtained for α, β and λ whatever the spell of time considered. For the sake of econometric estimation, one should retain any length of time τ allowing the economy to move forward to its target (its steady state on the corresponding sub-period) in a sufficiently clear way, while short enough not to imply too big a loss in terms of time series observations.

Choosing $\tau = 1$ - like CEL - makes sense if it can be assumed that convergence appears to take place significantly over a period of five years, and can therefore be subject to measurement on this length of time. $\tau = 2$ is otherwise a safer bet. For an estimation in levels, the latter is more appropriate, since inertia is a lot stronger for GDP in levels than it is for growth rates.

Following MRW, we assume in both cases that our economy converges between t-τ and t towards a steady-state that may be approximated by the average of the relevant variables on this spell of time[22]. A natural choice associated with τ for any given steady state variable x is then: $x^*_t = mean\left(x_{t-\tau}, \ldots, x_t\right)$.

---

[22] As the human capital variables are missing for the most recent period because they are defined in a recursive equation involving the future, averages for these variables are computed over all available dates.

Testing the Solow model

MRW and CEL rely on two methods in order to assess the consistency of the Solow model with the data.

As mentioned in section 1, they first use the estimates of the coefficients to compute values for α and β. These estimates are then compared to widely acknowledged values for these parameters. This can hardly be considered a statistical test of the model.

The prediction of the model that the sum of the coefficients of $\ln s_{it}^{k}$ and $\ln s_{it}^{h}$ should be the opposite of the coefficient on $\ln(n_{it} + g + \delta)$ (see equation (2)) provides a way of testing the model[23]. However, not all theoretical restrictions on estimated parameters are tested in this way, because the speed of convergence is in their model theoretically given by the expression:

$$\lambda = \left(1 - \alpha - \beta\right)\left(n + \delta\right)^{*}$$

MRW as well as CEL, following the tradition in this literature, sweep this problem aside by assuming a constant speed of convergence. Tradition however does not provide a test for the validity of this assumption. We attempt in this paper to test the additional theoretical restriction imposed on λ.

We estimate equation (2") under both assumptions of constant and endogenous speed of convergence. We provide tests of the adequation of the Solow model to the data in both cases.

A test of the Solow model under the assumption of constant speed of convergence

One may rewrite equation (2") as:

$$\ln y_{i,t} = \psi_0 \ln y_{i,t-\tau} + \psi_1 \ln\left(s_{i,t}^{k}\right)^{*} + \psi_2 \ln\left(s_{i,t}^{h}\right)^{*} + \psi_3 \ln(n_{i,t} + \delta)^{*}$$
$$+ \psi_4 \ln(n_{i,t} + \delta - s_{i,t}^{h})^{*} + \varepsilon_{it}$$

Estimating this model means inferring values for the structural parameters $\Xi = (\alpha, \beta, \lambda)$ from $\hat{\Psi} = \left(\hat{\psi}_0, \hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3, \hat{\psi}_4\right)'$, the vector of estimated parameters. Note that there are five estimated parameters for three parameters of interest. The model is therefore overidentified, and this provides a way of testing its validity that reaches beyond comparing the values obtained for α, β and λ to commonly admitted ones.

Let g be the non linear function relating the true vector of the reduced parameters to the structural parameters: $\Psi = g(\Xi)$. The latter can be obtained from the former as the solution to the following minimization program (Asymptotic Least Squares):

---

[23] This restriction on estimated parameters is rejected neither by MRW nor by CEL on the unrestricted augmented model. Whereas MRW obtain plausible values for technological parameters from the ex-ante restricted equation, CEL find unlikely values for α and β (the latter turning out to be significantly negative).

$$Min_{\Xi}\left(\hat{\Psi} - g(\Xi)\right)' \Omega\left(\hat{\Psi} - g(\Xi)\right)$$

The optimal choice for the weighting matrix Ω is the inverse of the covariance matrix of the unconstrained estimator, $V^{-1}(\hat{\Psi})$.

As :

$$N\left(\hat{\Psi} - g\left(\hat{\Xi}\right)\right)' \hat{V}_{\Psi}^{-1}\left(\hat{\Psi} - g\left(\hat{\Xi}\right)\right) \xrightarrow{L} \chi^2\left(\dim\Psi - \dim\Xi\right)$$

A test of the validity of the overidentifying constraints is provided by the comparison of the left hand side quantity to the critical value of a chi-squared with two degrees of freedom.

A test of the Solow model under the assumption of endogenous speed of convergence

Since $\lambda = (1-\alpha)(n+\delta)$, one would ideally estimate the following general specification, testing for the consistency of estimated coefficients as before:

$$\ln y_{i,t} = e^{\left(\varphi_0 + \varphi_1\left(n_{it} + \delta\right)\right)} \ln y_{i,t-\tau} +$$
$$\left[1 - e^{\left(\varphi_0 + \varphi_1\left(n_{it}+\delta\right)\right)}\right]\left[\theta_1 \ln\left(s^k_{i,t}\right)^* + \theta_2 \ln\left(s^h_{i,t}\right)^* + \theta_3 \ln(n_{i,t} + \delta)^* + \theta_4 \ln(n_{i,t} + \delta - s^h_{i,t})^*\right]$$
$$+ \varepsilon_{it}$$

where $\varphi_1 = -5\tau(1-\alpha)$

In particular, a test of the hypothesis $\varphi_1 = 0$ is a test of the validity of the traditional assumption of constant speed of convergence. Performing such a test however proves inconclusive, as estimations do not converge.

In order to obtain a more stable equation, we turn instead to a less demanding, more restricted test:

$$\ln y_{i,t} = e^{\left(\varphi_0 + \varphi_1\left(n_{it} + \delta\right)\right)} \ln y_{i,t-\tau} +$$
$$\left[1 - e^{\left(\varphi_0 + \varphi_1\left(n_{it}+\delta\right)\right)}\right]\left[\frac{\alpha}{1-\alpha} \ln\left(s^k_{i,t}\right)^* + \frac{\beta}{1-\alpha} \ln\left(s^h_{i,t}\right)^* - \frac{1}{1-\alpha} \ln(n_{i,t} + \delta)^* + \frac{1-\alpha-\beta}{1-\alpha} \ln(n_{i,t} + \delta - s^h_{i,t})^*\right]$$
$$+ \varepsilon_{i,t} \tag{3}$$

A test of the validity of the Solow model's prediction regarding the speed of convergence λ is then provided by testing the hypotheses $\varphi_0 = 0$ and $\varphi_1 = 5\tau(\alpha - 1)$ [24].

Imposing still more restrictions ex ante on equation (3) allows to distinguish five particular cases:

---

[24] Note that it is the rate of growth of population added to the assumed constant rate of depreciation which is multiplied by φ_1, so that the speed of convergence contains no constant term when $\varphi_0=0$.

A. original equation (3)

B. equation (3) with the restriction $\varphi_0 = 0$

C. equation (3) with the restriction $\varphi_1 = 0$

D. equation (3) with the restriction $\varphi_1 = 5\tau(\alpha - 1)$

E. equation (3) with the restrictions $\varphi_0 = 0$ and $\varphi_1 = 5\tau(\alpha - 1)$ (pure theoretical model).

Specifications B to E correspond to linear restrictions imposed on the coefficients of equation (3). They are therefore special cases of specification A.

One way of testing the validity of these restrictions consists in estimating specification A, and using ALS to check the consistency of estimated coefficients with the constraints imposed by specifications B to E. As for the case of constant speed of convergence, this is a test carried out ex post on estimated coefficients.

An alternative approach is to estimate all specifications A to E separately, using the same set of instruments, and use Sargan differences to test the restrictions imposed on the coefficients of the explanatory variables given this set of moment conditions: if the restrictions are valid, then the difference between the Sargan statistics computed from the restricted and unrestricted models follows a χ^2 with a number of degrees of freedom equal to the increase in the excess of moment conditions over the number of estimated parameters.

Results

Under the constant speed of convergence hypothesis: the Solow model is not rejected

We estimate equation (2") as explained in 2.2[25]. Table 1 sums up the results obtained[26] for various specifications.

For equations in levels, the set of instruments is made of the lagged values of all first differenced explanatory variables - except s_h[27]- from t-1 down to t-2. We instrument the first differenced equation by the lagged values of the same explanatory variables in levels from t-2 down to t-3. Both specifications are minimal in that they strike a compromise between estimation accuracy and economy of instruments (see Appendix for a discussion).

Table 1[28] - Conditional convergence (Standard errors in parentheses)

	OLS on levels (τ=2)	First differences instrumented by lagged levels (τ=1)	Levels instrumented by lagged first differences (τ=1)	Levels instrumented by lagged first differences (τ=2)
Constant	0.04 (0.37)	0.05 (0.01)	-0.00 (0.21)	2.41 (0.44)
Ln($Y_{-\tau}$)	0.93 (0.02)	0.82 (0.07)	0.93 (0.03)	0.65 (0.05)
Ln($n+\delta$)	-0.14 (0.13)	-0.07 (0.13)	-0.23 (0.07)	-0.40 (0.15)
Ln(s_k)	0.16 (0.03)	0.07 (0.05)	0.10 (0.02)	0.16 (0.06)
Ln(s_h)	0.01 (0.02)	-0.04 (0.02)	0.02 (0.01)	0.13 (0.04)
Ln($n+\delta$-s_h)	-0.01 (0.05)	0.02 (0.03)	0.05 (0.03)	0.20 (0.07)
Sargan (pvalue, df)		34.4 (0.06, 23)	36.4 (0.31, 33)	31.6 (0.07, 21)

OLS estimates are known to be trivially biased since the lagged endogenous variable includes the individual effect. One then expects the coefficient on lagged GDP to be biased upwards. Our results confirm this conjecture: the coefficient given by OLS is 0.93, to be compared with a value of 0.65 given by the GMM estimator for the same τ=2.

The choice of τ=2 for the first differenced model implies a severe loss in terms of number of time series observations. Besides, considering the low amount of autocorrelation on GDP growth rates, there is no need to extend the estimation period to ten years. Consequently, we only test the first differenced model under the assumption τ=1.

Standards errors are large, and few coefficients appear to be significantly different from zero.

---

[25] We use the Summers Heston dataset. Small countries whose population is less than a million, as well as oil producing countries are excluded from the sample. We end up with 78 countries after eliminating further countries for which human capital variables are not available.

[26] We use the program DPD98 written in Gauss for our linear GMM estimations (Arellano and Bond[1998]).

[27] Lagged values of lns_h are not included in the set of instruments because its compatibility with the rest of the instruments is rejected by difference Sargan tests. However, the same test leads to a strong acceptation of ln($n+d$-s_h) as an instrument.

[28] All GMM estimates displayed in this table are second step estimates. See Appendix for a discussion. Steady states variables are defined as the value corresponding to the middle of each convergence sub period.

GDP in level exhibits a strong time persistence for τ=1, pointing to a slow convergence towards steady state. The significance of the population growth rate and the saving rate appears to be good in both specifications. This is not the case however for the human capital variables, which turn out to be a lot more significant for the specification τ=2. Indeed, it makes sense to consider that human capital investment does not have a sizeable impact on GDP until roughly a decade. Note that neither our data nor our model allow for a possible heterogeneity in the efficiency of the schooling systems.

These considerations lead us to retain the specification τ=2 for the equation written in levels.

The fact that we have a larger number of estimated coefficients than structural parameters implies that the validity of the model may be tested through the consistency of these estimates.

It turns out that the p-value for the χ^2 test presented above is 75%. The ALS yield the following result for the minimal set of instruments:

parameter	ALS estimate
$e^{-10\lambda}$	0.66 (0.04)
λ	0.04
α	0.30 (0.07)
β	0.26 (0.03)

λ = 0.04 implies a half life of conditional convergence equal to 15 years.

We obtain a realistic value of 0.30 for α. The estimate for β implies a share of skilled work in aggregate labour remuneration of roughly 40%. Interpreting this figure is not straightforward considering the heterogeneity of our sample, which includes both developing and OECD countries. Returns to capital are theoretically higher in developing countries because capital is less abundant. Relative return to skill may also be higher there as this category of labour is scarce. Aggregate factor shares in national income may nevertheless be viewed as roughly constant across the sample as the opposite holds for developed countries: capital and skill are more abundant and their relative pay is lower.

Under endogenous speed of convergence

The linear case presented above is not consistent with the Solow model, from which one derives an endogenous speed of convergence. We now turn to this issue, following the methodology presented in section 3.2.

We start by estimating the most general specification A corresponding to equation (3). We then carry out ALS to test the constraints imposed by specifications B to E.

Table 2 - Tests ex post of the theoretical constraints based on the unrestricted specification
(Standard errors in parentheses)

	A (Benchmark)	B $\varphi_0 = 0$	C $\varphi_1 = 0$	D $\varphi_1 = 10(\alpha-1)$	E $\varphi_0 = 0$ $\varphi_1 = 10(\alpha-1)$
α	0.31 (0.06)	0.33 (0.06)	0.26 (0.05)	0.41 (0.04)	0.40 (0.04)
β	0.26 (0.08)	0.18 (0.07)	0.37 (0.06)	0.12 (0.06)	0.11 (0.06)
φ_0	-0.32 (0.15)		-0.51 (0.11)	-0.11 (0.12)	
φ_1	-2.68 (-1.32)	-4.50 (1.02)			
χ^2	15.46	4.80	4.14	6.65	7.41
DF	28	1	1	1	2
P-value	0.97	0.03	0.04	0.01	0.02

ALS lead to the rejection of all constraints, in particular of the one based on a constant speed of convergence (B), as well as the pure Solow theoretical model (E). It seems from these results that the Solow model is rejected by the data when it is taken seriously, that is when all the theoretical restrictions it imposes are taken into account. Note however that this conclusion may be due to the choice of the reference specification on which ALS are performed. Recall that specification A is not the most general, as explained in section 3.2. The choice made is therefore arbitrary, and gives a particular role to specification A.

We therefore turn to the alternative method based on Sargan differences. Restrictions are now imposed ex ante and tests are carried out on the corresponding estimations. Table 3 presents the results.

Table 3 - Tests based on the estimation of restricted specifications
(Standard errors in parentheses)

	A (Benchmark)	B $\varphi_0 = 0$	C $\varphi_1 = 0$	D $\varphi_1 = 10(\alpha-1)$	E $\varphi_0 = 0$ $\varphi_1 = 10(\alpha-1)$
α	0.31 (0.06)	0.40 (0.05)	0.28 (0.06)	0.38 (0.05)	0.40 (0.04)
β	0.26 (0.08)	0.25 (0.05)	0.35 (0.07)	0.18 (0.07)	0.19 (0.04)
φ_0	-0.32 (0.15)		-0.55 (0.10)	-0.19 (0.14)	
φ_1	-2.68 (-1.32)	-3.86 (0.89)			
Sargan	15.46	28.55	18.75	14.5	25.96
Sargan based on W.M.(E)	22.92	23.51		24.41	25.96
DF	28	29	29	29	30
P-value	0.97	0.49	0.93	0.99	0.68

These specifications can be distinguished by the constraint they impose on φ_0 and φ_1. In each case, estimates of α and β are significant and consistent with each other given standard errors.

The estimates of φ_1 obtained from specification (A) and (B) are consistent with the values computed from the corresponding estimates of α and the theoretical relationship $\varphi_1 = 10(\alpha-1)$.

The Sargan difference tests validate the restrictions that specification (D) imposes on (A) and (E) on (B). One problem in implementing this test is due to the fact that the difference of Sargan statistics are negative. We deal with this by carrying out the test on Sargan statistics obtained by using the same second step weighting matrix. Conclusions are unchanged.

Turning to tests of the restriction $\varphi_0 = 0$:

The difference tests based on Sargan statistics obtained by using the same second step weighting matrix (that of specification (E)) do not reject the restriction $\varphi_0 = 0$. Note however that the difference tests obtained by using different weighting matrices (second step matrices obtained from respectively specifications (D) and (E)) do not confirm this result. While using a unique weighting matrix seems to us to more relevant, one has to admit that contrary to the previous case, conclusions are not robust to a change in the way the test is carried out. They must as a result be considered as fragile.

Despite the robustness of the estimates for α and β across the specifications, and despite the validation of the constraint $\varphi_1 = 10(\alpha - 1)$, the fragile validation of $\varphi_0 = 0$ does not allow to conclude sharply in favour of the Solow model.

This is all the more obvious considering the results given by the first method presented above. Whether this fragility points to an insufficient number of observations or must be interpreted as a sign of the weakness of the Solow model as such may not be concluded from this study.

Conclusion

Two main conclusions may be drawn from this paper.

First, the estimations carried out under the assumption of constant speed of convergence show that the estimation strategy we suggest, based on the estimation of the model in levels instrumented by lagged first differences, is both more rigorous than MRW's and more efficient than CEL's. It is true that these results were obtained at the cost of an additional identifying assumption, that of constant correlation between regressors and the individual effect (stationarity assumption). We do not believe this assumption to be strong. We are comforted in this by the fact that Sargan tests do not reject the consistency of the orthogonality conditions generated by this specification. Asymptotic least squares applied to our linear GMM estimator confirms both the robustness of our estimates and the compatibility of the constraints imposed on them by the theoretical model. Besides, the estimated value for GDP elasticity to physical capital is close to 0.30, a value generally deemed to be realistic.

Secondly, this methodological improvement is not sufficient in itself to provide a validation of the Solow model for two reasons.

To begin with, one may not claim to carry out a rigorous test of a model by estimating an equation obtained by imposing a restriction not derived from the model itself. From this point of view, assuming a constant speed of convergence across countries is theoretically inconsistent. It turns out that our tests carried out under the more general assumption of endogenous speed of convergence do not validate the restriction that the speed of convergence is constant: even if the constant term included in λ appears to be significant, so is the endogenous one. The assumption of constant λ must therefore be viewed at best as a poor approximation only weakly supported by the data.

Given this, tests of the rigorous model where the expression of λ is derived from the Solow model itself, show that the theoretically consistent convergence equation is not fully consistent with the data either. Note however that the estimates of the elasticities of output to physical and human capital are hardly affected, compared to the more restricted case of constant λ.

These results obviously lead to question the adequation of our theoretical framework with the data. This may point either to a misspecification of the production function, to the existence of strong heterogeneity across countries or to the inadequation of the Solow framework altogether. Further research is obviously needed to draw a clear conclusion on the implications in terms of test of the Solow model.

References

ARELLANO Manuel, BOND Stephen, « Some Tests of Specification for Panel Data: Monte Carlo Evidence and an application to Employment Equations », *Review of Economic Studies*, 58, 1991

ARELLANO Manuel, BOND Stephen, « Dynamic Panel Data Estimation Using DPD98 for Gauss: A Guide for Users », 1998

ARELLANO Manuel, BOVER Olympia, « Another Look at the Instrumental Variable Estimation of Error-Components Models », *Journal of Econometrics*, 68, 1995

BARRO Robert, LEE Jong-Wha, « International Comparisons of Educational Attainment », *Journal of Monetary Economics*, 32, 1993

BLUNDELL Richard, BOND Stephen, « Initial Conditions and moment restrictions in Dynamic Panel Data Models », *Journal of Econometrics*, 87, 1998

CASELLI Francesco, ESQUIVEL Gerardo, LEFORT Fernando, « Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics », *Journal of Economic Growth*, 1:3, 1996

ISLAM, Nazrul, « Growth Empirics: A Panel Data Approach », *Quarterly Journal of Economics*, 110, 1995

KNIGHT Malcolm, LOAYZA Norman, VILLANEUVA Delano, « Testing the Neoclassical Growth Model », *IMF Staff Papers*, 40, 1993

MANKIW Gregory, ROMER David, WEIL David, « A Contribution to the Empirics of Economic Growth », *Quarterly Journal of Economics*, 107:2, 1992

QUAH Dany T., « Empirics for growth and distribution: stratification, polarization and convergence clubs », *Journal of Economic Growth,* 1997

SOLOW Robert, « A Contribution to the Theory of Economic Growth », *Quarterly Journal of Economics*, 70, 1956

SUMMERS Robert, HESTON Alan, « A new set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries », *Review of Income and Wealth*, 34, 1988

TEMPLE Jonathan, « The New Growth Evidence », *Journal of Economic Literature*, 37, 1999

Appendix

Solving the model

The production function is written:

$$Y(t) = A(t)K(t)^{\alpha} H(t)^{\beta} L(t)^{1-\alpha-\beta}$$

Labour (which equals population by assumption) and technical change (Hicks neutral) are supposed to grow at exogeneous and constant rates *n* and *g:*

$$N(t) = N(0)e^{nt}$$
$$A(t) = A(0)e^{gt}$$

$$N(t) = H(t) + L(t)$$

Values per unit of labour (or population i.e. per head) are defined as:

$$y = \frac{Y}{N} \quad k = \frac{K}{N} \quad h = \frac{H}{N} \quad l = \frac{L}{N}$$

Hence, rewriting the production function:

$$y(t) = A(t)k(t)^{\alpha} h(t)^{\beta} l(t)^{1-\alpha-\beta}$$

Accumulation equations, and steady state

The physical capital accumulation equation may be written:

$$\dot{K} = s_k Y - \delta_K K$$

$$\dot{k} = s_k y - (n + \delta_K)k$$

Human capital accumulation is given by:

$$\dot{H} = s_h(H + L) - \delta_H H$$

$$\dot{h} = s_h - (n + \delta_H)h$$

The dynamics of unskilled labour accumulation is then obtained from the relation $l + h = 1$:

$$\dot{l} = -\dot{h}$$

In order to define a limit trajectory, we make assumptions on the growth rates of per head variables when this steady state is reached. We assume that the ratio of output to the stock of physical capital is fairly constant over time, therefore that the rates of growth of per head output and per head capital must be identical at steady state. We do not

however assume that the ratio of the skilled workforce to output is constant, in other words that the productivity of skilled workers does not grow.

Log-differentiating the production function yields:

$$\frac{\dot{y}}{y} = \alpha\frac{\dot{k}}{k} + \beta\frac{\dot{h}}{h} + (1-\alpha-\beta)\frac{\dot{l}}{l} + g$$

Hence, equating the growth rates of y and k yields:

$$\left.\frac{\dot{y}}{y}\right|_* = \left.\frac{\dot{k}}{k}\right|_* = \frac{g}{1-\alpha}$$

Note that output and physical capital per head do not grow at the same rate as human capital and unskilled labour on the limit trajectory. In order to derive the convergence path towards the limit trajectory by using linear approximation around this trajectory, we make the following assumption as to the form of the solution we seek to obtain for per head variables:

$$x = \tilde{x}e^{r_x t}$$

where r_x is the rate of growth of per head variable x on the limit trajectory and $\tilde{x}$ is by definition constant on the limit trajectory.

Note that h and l are unchanged by this transformation, as their rates of growth are zero on the limit trajectory. However, y and k are affected. Let us then rewrite the accumulation equation for physical capital as well as the production function in terms of the transformed variables.

$$\dot{\tilde{k}} = s_k\tilde{y} - (n+\delta_K + \frac{g}{1-\alpha})\tilde{k}$$

$$\tilde{y} = \tilde{k}^\alpha\tilde{h}^\beta\tilde{l}^{1-\alpha-\beta}$$

We shall from this point omit the tildas on the transformed variables for the sake of simplicity.

The steady state is then given by:

$$\frac{s_k y^*}{k^*} = n+\delta_K + \frac{g}{1-\alpha}$$

$$h^* = \frac{s_h}{n+\delta_H}$$

$$l^* = 1 - \frac{s_h}{n+\delta_H}$$

Conditional convergence towards the limit trajectory

Let us first approximate the accumulation equations around the steady state defined above in terms of transformed variables, starting with physical capital:

$$\dot{k} = \frac{\partial \dot{k}(k,h,l)}{\partial k}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}}(k-k^*) + \frac{\partial \dot{k}(k,h,l)}{\partial h}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}}(h-h^*) + \frac{\partial \dot{k}(k,h,l)}{\partial l}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}}(l-l^*)$$

Obviously:

$$\frac{\partial \dot{k}(k,h,l)}{\partial k}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = s_k \frac{\partial y(k,h,l)}{\partial k}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} - (n+\delta_K + \frac{g}{1-\alpha})$$

By definition of α :

$$\alpha = \frac{\partial y(k,h,l)}{\partial k}\frac{k}{y}$$

Hence:

$$\frac{\partial \dot{k}(k,h,l)}{\partial k}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = s_k \frac{\alpha y(k^*,h^*,l^*)}{k^*} - (n+\delta_K + \frac{g}{1-\alpha})$$

By definition of the steady state

$$\frac{\partial \dot{k}(k,h,l)}{\partial k}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = (\alpha-1)(n+\delta_K + \frac{g}{1-\alpha})$$

One derives in the same way:

$$\frac{\partial \dot{k}(k,h,l)}{\partial h}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = s_k \frac{\partial y(k,h,l)}{\partial h}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = s_k \frac{\beta y(k^*,h^*,l^*)}{h^*} = \frac{\beta k^*}{h^*}\frac{s_k y(k^*,h^*,l^*)}{k^*}$$

Hence :

$$\frac{\partial \dot{k}(k,h,l)}{\partial h}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = \frac{\beta k^*}{h^*}\left[n+\delta_K + \frac{g}{1-\alpha}\right]$$

Similarly:

$$\frac{\partial \dot{k}(k,h,l)}{\partial l}\bigg|_{\substack{k=k^* \\ h=h^* \\ l=l^*}} = \frac{(1-\alpha-\beta)k^*}{l^*}\left[n+\delta_K + \frac{g}{1-\alpha}\right]$$

Finally:

$$\dot{k} = (n + \delta_K + \frac{g}{1-\alpha})\left[(\alpha-1)(k-k^*) + \frac{\beta k^*}{h^*}(h-h^*) + \frac{(1-\alpha-\beta)k^*}{l^*}(l-l^*)\right]$$

$$\dot{h} = -\left(n + \delta_H\right)\left(h - h^*\right)$$

$$\dot{l} = -\left(n + \delta_H\right)\left(l - l^*\right)$$

Making the first order approximation at the vicinity of the steady state:

$$y = y^*, k = k^*, h = h^*, l = l^*$$

Plugging the previous approximations of the dynamics of factor accumulation into the log-differentiated per head production function yields:

$$\frac{\dot{y}}{y} = \alpha(n + \delta_K + \frac{g}{1-\alpha})\left[(\alpha-1)\frac{k-k^*}{k^*} + \beta\frac{h-h^*}{h^*} + (1-\alpha-\beta)\frac{l-l^*}{l^*}\right]$$

$$- (n+\delta_H)\beta\left(\frac{h-h^*}{h^*}\right) - (n+\delta_H)(1-\alpha-\beta)\left(\frac{l-l^*}{l^*}\right)$$

In order to factorize the analogous expression obtained from the traditional augmented Solow model, the previous literature on convergence (for instance MRW) assumes that depreciation rates on physical and human capital are identical. For the same reason, we have to make the assumption that the rates of depreciation are identical modulo the limit trajectory common rate of growth of output per head and physical capital per head:

$$\delta_H = \delta_K + \frac{g}{1-\alpha} = \delta$$

This restriction is obviously arbitrary. Note however that the rates of depreciation involved are unknown and that the previous condition is by no means more restrictive than the one imposed by MRW. Indeed, this alone does not provide a justification. The assumption is obviously made for convenience as it allows to obtain a simple representation of the dynamics of transformed output. The latter must consequently be viewed as a simplification of the « true » convergence path.

The previous expression may then be factorized into:

$$\frac{\dot{y}}{y} = (n+\delta)(\alpha-1)\left[\alpha\frac{k-k^*}{k^*} + \beta\frac{h-h^*}{h^*} + (1-\alpha-\beta)\frac{l-l^*}{l^*}\right]$$

Since one may also write around the steady state:

$$\frac{y-y^*}{y^*} = \alpha\frac{k-k^*}{k^*} + \beta\frac{h-h^*}{h^*} + (1-\alpha-\beta)\frac{l-l^*}{l^*}$$

Factorising the latter expression into the former yields:

$$\frac{\dot{y}}{y} = (n+\delta)(\alpha-1)\frac{y-y^*}{y^*}$$

Or alternatively:

$$\frac{d \ln y}{dt} = \lambda(\ln y^* - \ln y)$$

$$\lambda = (1-\alpha)(n+\delta)$$

Solving the differential equation and writing the limit conditions:

$$\frac{d\left[\ln y - \ln y^*\right]}{dt} = -\lambda(\ln y - \ln y^*)$$

$$\ln \frac{y}{y^*} = Ce^{-\lambda t}$$

$$C = \ln y_0 - \ln y^*$$

The solution is therefore given by:

$$\ln y = e^{-\lambda t} \ln y_0 + \left(1 - e^{-\lambda t}\right) \ln y^*$$

Recall that the above equation has been derived using the concept of transformed output (i.e. such that the rate of growth on the limit trajectory is zero). We seek however to estimate a convergence equation involving the initial measurable per head variables. Obviously, the previous equation is unchanged when expressed in terms of per head variables.

Estimating the previous equation requires that the expression of the steady state is explicited. By definition (in per head variables):

$$\ln y^* = \alpha \ln k^* + \beta \ln h^* + (1-\alpha-\beta)\ln l^* + \ln A^*$$

$$\ln k^* = \ln y^* + \ln s_k - \ln(n+\delta)$$

$$\ln h^* = \ln s_h - \ln(n+\delta)$$

$$\ln l^* = \ln\left(1 - \frac{s_h}{n+\delta}\right) = \ln(n+\delta-s_h) - \ln(n+\delta)$$

Hence, returning to per unit of efficient labour values:

$$\boxed{\ln y^* = \frac{\alpha}{1-\alpha} \ln s_k + \frac{\beta}{1-\alpha} \ln s_h + \frac{1-\alpha-\beta}{1-\alpha} \ln(n+\delta-s_h) - \frac{1}{1-\alpha} \ln(n+\delta)}$$

Hence:

$$\ln y = e^{-\lambda t} \ln y_0 + \left(1 - e^{-\lambda t}\right)\frac{\alpha}{1-\alpha} \ln s_k + \left(1 - e^{-\lambda t}\right)\frac{\beta}{1-\alpha} \ln s_h$$

$$+ \left(1 - e^{-\lambda t}\right)\frac{1-\alpha-\beta}{1-\alpha} \ln(n+\delta-s_h) - \left(1 - e^{-\lambda t}\right)\frac{1}{1-\alpha} \ln(n+\delta)$$

The terms appearing in this equation and involving either the level or the rate of growth of technology (i.e. g or A) are unobserved and included in the residual.

By definition of y:

$$\ln y_t = \ln \frac{Y_t}{N_t} - \ln A_t = \ln \frac{Y_t}{N_t} - \ln A_0 - gt$$

The idiosyncratic level of technology as well as the common trend of technical change are therefore included in the residual.

One can write this equation in the general way for every sub-period t of length 5 years:

$$\ln\left(\frac{Y_t}{L_t}\right) = e^{-5\lambda\tau}\ln\left(\frac{Y_{t-\tau}}{L_{t-\tau}}\right) + \left(1 - e^{-5\lambda\tau}\right)\frac{\alpha}{1-\alpha}\ln s^{k*}_{t-\tau} + \left(1 - e^{-\lambda\tau}\right)\frac{\beta}{1-\alpha}\ln s^{h*}_{t-\tau}$$
$$+ \left(1 - e^{-5\lambda\tau}\right)\frac{1-\alpha-\beta}{1-\alpha}\ln\left(n + \delta - s_h\right)^*_{t-\tau} - \left(1 - e^{-5\lambda\tau}\right)\frac{1}{1-\alpha}\ln\left(n + \delta\right)^*_{t-\tau} + \varepsilon_t$$

with the additional relationship

$$\lambda = \left(1 - \alpha\right)\left(n + \delta\right)^*$$

and the residual specified as:

$$\varepsilon_{it} = u_i + \eta_t + v_{it}$$

First step vs two step estimates

All results displayed in this paper are second step GMM estimates. The first step estimator is known to be consistent, yet sub-optimal as it weighs orthogonality conditions[29] regardless of their precision. However, under small sample conditions, the convergence of the second step weighting matrix is likely to be only partial. We therefore systematically check that:

- The second step estimates remain close to the values obtained from the first step.

- The second step weighting matrix is not ill-conditioned, making the inversion process somewhat elusive (as a matter of fact, the ratio of the largest eigen value to the smallest turns out to grow with the size of the instrument set, but never in a way that makes the process of inversion numerically insignificant).

What one should expect from comparing the first step estimator to the second step estimator, is an increase in precision (smaller standard errors) but small variations in the coefficient estimates. This is indeed what we observe (see Table 3), which leads us to retain the second step estimator.

---

[29] the weighting matrix is the identity matrix.

Table 4 - GMM on the model in levels[30], τ=2 (Standard errors in parentheses)

Weighting matrix	First step (I)	Second step (V^1)
Constant	1.61 (0.74)	2.41 (0.44)
Ln($Y_{-\tau}$)	0.68 (0.07)	0.65 (0.05)
Ln($n+\delta$)	-0.53 (0.32)	-0.40 (0.15)
Ln(s_k)	0.26 (0.08)	0.16 (0.06)
Ln(s_h)	0.13 (0.06)	0.13 (0.04)
Ln($n+\delta$-s_h)	0.18 (0.14)	0.20 (0.07)

Number of instruments

In Table 1, the chosen set of instruments for the model in levels is incomplete since not all possible instruments are included (i.e. more lagged values could be added up to four in all). Before asserting that the corresponding GMM estimator is sub-optimal, one should bear in mind that we are dealing with a small sample, and that making use of all possible orthogonality conditions does not necessarily make sense: a larger set of instruments implies a bigger variance matrix of orthogonality conditions, which is more difficult to invert (see preceding Appendix).

To deal with this difficulty, we analyze the response of our estimates to increases in the size of the instrument set. The results obtained for the full set of instruments (consistent with the assumption of weak exogeneity) are shown in Table 5.

Table 5 - The model in levels instrumented by the first differences of explanatory variables lagged up to four times (Standard errors in parentheses)

Constant	1.92 (0.37)
Ln($Y_{-\tau}$)	0.68 (0.04)
Ln($n+ \delta$)	-0.50 (0.15)
Ln(s_k)	0.18 (0.04)
Ln(s_h)	0.13 (0.03)
Ln($n+\delta$-s_h)	0.19 (0.05)
Sargan (pvalue, df)	37.5 (0.27, 33)

The difference Sargan test does not lead to the rejection of the additional instruments. What is more, the estimation remains close to the one obtained using the smaller set of instruments, with a significant reduction in the standard errors[31]. The latter should not be viewed with excessive optimism, considering the small size of the sample. Our consistency criteria should however be considered as fulfilled. These results do not

---

[30] As in section 4.1., instruments are first differences of the explanatory and the endogenous variables lagged once and twice.

[31] One should bear in mind that small samples entail the risk of underestimating second step standard errors. Yet, the observed reduction in standard errors between the two steps is not sharp enough to raise suspicion.

invalidate the use of all instruments compatible with the hypothesis of weak exogeneity, but they do not rest on it.