Adriana Di Liberto
Università di Sassari, University College London and CRENoS
Uctpadi@ucl.ac.uk

James Symons
University College London
j.symons@ucl.ac.uk

SOME ECONOMETRIC ISSUES IN CONVERGENCE REGRESSIONS

Abstract: Despite the abundance of different econometric techniques introduced in the empirical literature on convergence, it is usually assumed that shocks are uncorrelated across countries. This is surely unlikely for most of the datasets considered and we investigate a possibility so far ignored, namely the annual panel estimator where shocks are allowed to be correlated across countries. Our analysis is restricted to the case of more time periods than countries (T>N) which allows us to estimate by Maximum Likelihood with an unrestricted variance-covariance matrix of cross-country shocks. The paper examines by Monte Carlo robustness against certain possible mis-specifications, namely measurement error and heterogeneity of the convergence coefficients. Our analysis indicates that ML estimators are robust to plausible measurement error and variation of convergence rates across countries and are more efficient than conventional estimators for plausible values of cross-country error correlation. We consider in detail the relationship between the distribution of the ML estimator and the initial conditions. Applying our findings to a panel of OECD countries for the post-war period, we show that ML is effectively unbiased and more efficient than or conventional panel estimators OLS on a cross-section of countries. We argue the reason this estimators is so well behaved is that many OECD countries were far from their equilibrium values at the beginning of the period.

Financial support provided by MURST (Research Project on “European Integration and regional disparities in Economic Growth and Unemployment”) is gratefully acknowledged.

January 1999

1. Introduction

Baumol (1986) and Barro and Sala-i-Martin (1995) have investigated the convergence hypothesis by regressing growth in a collection of regions over certain time intervals on the initial level of GDP/head. The equation that is estimated - call it the *Baumol*-regression (B-regression)- is usually deduced from a certain autoregressive process implied by neo-classical growth theory. The B–regression (OLS in a cross-section of countries) was at first considered as the benchmark in empirical studies on growth and convergence. More recently an increasing number of studies have tryied to exploit both the cross-section and the time series nature of the dataset. There is now a huge literature on this topic[1] and a number of econometric techniques have emerged. It is natural to ask which of these and other methods is the best way to estimate convergence rates. This is important because one is interested in inferring from the estimate how rapidly countries or regions will converge and even tiny differences in the estimated coefficient imply enormous differences in the predicted patterns of convergence.

In this paper we investigate a possibility so far ignored, namely the annual panel estimator where shocks are allowed to be correlated across countries. This is the Maximum Likelihood (ML) estimator in these circumstances. Other authors[2] have used panel methods but have assumed effectively that shocks are uncorrelated across countries, surely wrong for most of the datasets considered. In principle, panel data approaches exploit more data than the B–regression (and Barro and Sala-i-Martin's subsequent development of this technique) and hence might be expected to be more efficient. Against this, the B–regression is likely to be more robust against certain possible mis-specifications. There have been some suggestions that panel estimates are incorregibly biased[3] together with recent arguments in favour of the B–regression[4]. Indeed, we show below that estimating the B–regression is more efficient than maximum likelihood on the full panel, provided that shocks are not too correlated across regions or countries. For cross correlations of the order that arise in OECD countries, however, ML dominates the B–regression. It is also

---

[1] See e.g. the survey by Durlauf and Quah (1998).
[2] e.g. Knight *et al.* (1993), Islam (1995), Caselli *et al.* (1997) *inter alia.*
[3] See Pesaran *et al.* (1997).
[4] See Shioji (1998).

true that ML is more efficient than these panel methods which do not allow for correlations in cross-country shocks. Ignoring cross-sectional correlation leads not only to efficiency losses: it means also that inference is distorted, given that, as is easily demonstrated, important cross-correlations exist in OECD data. It is fair to say that there is hardly a consistently estimated standard error in this entire literature. The paper examines these issues by Monte Carlo. We apply our findings to a panel of OECD countries. On balance we find some evidence against convergence. Our analysis will be restricted to the case that there are more time periods than countries (T>N) which allows us to estimate an unrestricted variance-covariance matrix of cross-country shocks. In the opposite case, N>T, some restriction of the covariance matrix is necessary for the analysis. We do not discuss this case.

Our main contribution is to show that, for data-sets such as the OECD, ML is effectively unbiased and more efficient than the B–regression or conventional panel estimators. Our analysis indicates moreover that both the B–regression and ML estimates are robust to plausible measurement error and variation of convergence rates across countries. We show the reason these estimators are so well behaved is that many OECD countries were far from their equilibrium values in 1950. Our contribution here is consider in detail the relationship between the distribution of the ML estimator and the initial conditions[5]. We show that, unless initial conditions are sufficiently extreme, confidence intervals for the convergence rate are unsatisfactorily wide. We offer a likelihood ratio test of the fixed-effects model (a general test of the convergence hypothesis), and show why, in small samples, the true size of the test is much lower than nominal levels. We construct tests of correct size by Monte Carlo. This analysis is similar in some respects to a test proposed by Evans and Karras (1996), except that we allow for a general cross-sectional covariance matrix.

2. Econometric Issues in Convergence Regressions: testing absolute $\beta-$convergence

In this paper we analyse unconditional convergence, having in mind the OECD economies or the provinces of a given country. The general point that efficiency and inference are improved by exploiting the cross-

---

[5] Shioji (1998) has considered the importance of the initial conditions in bias reduction.

sectional correlations of groups of economies applies also to studies of conditional convergence and it may well be that our findings provide a useful direction for future work in this area. Thus, the convergence discussed in this paper is, in the terminology of Barro and Sala i Martin (op. cit.), absolute β–convergence. Formally, we say a set of countries exhibits absolute β–convergence if, for all pairs i, j in the set, and at all times t

$$\lim_{r} E_{t}(y_{it+r} - y_{jt+r}) = 0$$

where y_{it} and y_{jt} are the logarithms of output in countries *i* and *j* respectively. The conclusion of Barro and Sala i Martin is that absolute convergence tends to exist for homogenous regions (the states of the U.S., Japanese Prefectures, European regions) but not for more heterogeneous units (the world as a whole). Barro (e.g. 1998) concludes that convergence for this larger set is apparent only when one controls for a number of variables (conditional convergence).

Assume that the logarithm of output per capita in country i evolves according to

$$(1) \qquad (d/dt)(y_{i} - y_{\bullet}) = -\beta(y_{i} - y_{\bullet})$$

where y. refers to output per-capita in a perhaps hypothetical leading country. Equation (1) can be derived as a log-linear approximation around steady state of a single-good neo-classical growth model with labour-augmenting technical progress[6]. In this approach, the parameter β and the GDP steady state, y., depend on depreciation rates, savings rates and the rates of growth of population and technology. In a more general framework, the first three are endogenous variables, themselves functions of output per capita. We shall interpret (1) as a reduced form from such a model. Equation (1) is consistent also with other models in which convergence is not due to physical or human capital accumulation but to technology transfer.

To estimate β equation (1) is replaced by

---

[6] See Barro and Sala-i-Martin, *op. cit*.

(2) $$(y_{it} - y_{it-1}) - (y_{\bullet t} - y_{\bullet t-1}) = -\beta(y_{it-1} - y_{\bullet t-1}) + u_{it}$$

where $u_t = (u_{it},...,u_{nt})'$ is a vector white noise process i.e. the u_{it} are serially independent but perhaps contemporaneously correlated.

Equation (1) is seldom estimated as it stands. Rather, by repeated substitution one obtains

(3) $$y_{iT} - y_{i0} = -[1-(1-\beta)^T](y_{i0} - y_{\bullet}) + u_i$$

where

$$u_i = u_{iT} + (1-\beta)u_{iT-1} + ... + (1-\beta)^{T-1}u_{i1}$$

This can be estimated for a sample of countries and an estimate of β recovered. Baumol (1986) appears to have been the first to estimate β in this way. Barro and Sala-i-Martin (1995) have made much use of it. We shall call it the B-regression.

To estimate β, since the $y_{.t}$ process is unobserved, we choose at random a comparator country *j* and replace (2) by

(4) $$y^*_{it} - y^*_{it-1} = -\beta y^*_{it-1} + v_{it}$$

where y^*_{it} is output in i less output in *j* and $v_{it} = u_{it} - u_{jt}$. Note that absolute β–convergence, as we have defined it, amounts to asserting β>0. If N<T, (3) can be estimated by ML techniques with an unrestricted covariance matrix, allowing shocks to be correlated across countries, which is almost certain to be true. Expressing the data as deviation from a comparator country will eliminate some of the cross-sectional error correlation but not all unless:

(5) $$v_{it} = \eta_t + \varepsilon_{it}$$

where η_t is a time-varying common stochastic component[7]. This could not hold when international shocks impinge differently on different countries.

2.1. Possible Mispecifications

The B–regression has a number of things in its favour. First, the method has no difficulties with N > T whereas an ML attack on (4) will encounter problems in this case[8]. Second, as we shall see below, the method is robust to a number of plausible specification errors[9]. The obvious weakness with the approach is that it is a non-standard method of estimating the family of time-series given by (4) and is hardly likely to be efficient. A development of this method, presumably to increase efficiency, is to construct observations of (3) for shorter time periods, decades say. A related method (Barro, 1998) is to average (4) over time periods and estimate using the initial value as an instrument for the period-average GDP.

Using the ML approach, we shall regard the vector process y_t as having commenced at t=0 and conduct inference conditional on y_o. We do not regard the data as a realisation of a stationary vector process commencing in the distant past. For our sample, and in most samples in this literature, the presence of undeveloped regions which are many standard deviations from the assumed long-run mean makes this assumption essential. Thus many of the i processes in (4) are not even approximately realisations of a stationary time-series but are best thought of as exponential decay. As it happens, as we shall see below, this leads to increased efficiency in estimation.

The problem with estimating (4) is that one is trying to obtain precise estimates of the autoregressive parameter of a process close to the unit circle. Usually this is tricky because the estimator is badly behaved in small samples. For example, in the univariate case, the OLS estimator $\hat{\rho}$ of

---

[7] As assumed by Pesaran *et al.* (1997).

[8] A possible argument against the adoption of a SUR approach is that, assuming series are stationary, the SUR estimator has low power unless N is appreciably less than T, while it does not even exist when N≥T. See Evans and Karras (1996) on this point.

[9] Moreover, since the LHS can be rescaled to give annual growth and the RHS can be interpreted as "GDP-gap", one can obtain an instructive graph of growth against gap.

(6) $$z_t = \rho z_{t-1} + \varepsilon_t \qquad (t = 1,...T,\ z_o \text{ fixed})$$

is biased down by $2\rho/T$ for $\rho < 1$[10]. This is a large T result. For given moderate T and ρ close to unity, the distribution of $\hat{\rho}$ will be indistinguishable from the Dickey-Fuller distribution (for which this bias formula is a little different). What is mitigating in the cases we have in mind is that considering a panel of countries reduces bias considerably. Thus with, for example, $N\approx20$ and $T\approx40$ we have effectively 800 observations so that such bias is of the order of .002, non-trivial but unimportant in context. One should note however that if included regressors have genuinely different parameters then the false imposition of equality in estimation can bias β in the opposite direction i.e. towards 0[11]. It is likely that the B–regression (4) is more robust against this mis-specification.

A second mitigation is the initial conditions: z_0 in (6) or the set y^*_{i0} in (4). The bias estimates discussed above are for large T and the contribution of the initial conditions is of lower order than $1/T$. However, for fixed T, the distribution of $\hat{\rho}$ in (6) concentrates on ρ with variance proportional to σ^2/z_0^2 as z_o grows[12]. Thus the relevant parameter is the initial condition measured in units of the innovation standard deviation and, for typical samples, values of this can be 40 or more. In these circumstances the estimate will be accurate and precisely determined. Thus the small sample properties of the ML estimator of β are subject to two conflicting influences : the closeness to the unit circle induces the distribution to behave like a multivariate Dickey-Fuller distribution, while the extreme initial conditions induce normality. In such circumstances, Monte Carlo seems the only way to derive properties of the distribution.

Finally one might want to test H_0: $\beta = 0$ in (4). In these circumstances, (4) is a family of random walks and a version of the Dickey-Fuller test would need to be performed. If one took seriously the possibility that the

---

[10] See Grubb and Symons (1987) for a general discussion.

[11] Robertson and Symons (1991) discuss this possibility. In the event that the forcing variable is a random walk, the false imposition of parameter equality across countries will produce an estimate $\beta=0$ in large samples.

[12] Evans and Savin (1983).

v_{it} are correlated across i, critical values could be calculated by Monte Carlo, wherein the artificial random vector v_t would be selected with contemporaneous covariance matrix given by the sample covariance matrix of $y^*_{it} - y^*_{it-1}$. In the case we shall study, and in like works, one is certain to reject the null, because under H_o:β=0, the β–regression (4) should return an estimate on y^*_{io} of zero and the graph of growth against gap should reveal no systematic relationship. Neither is remotely the case.

3. ML versus the B–Regression and Pooling: a Monte Carlo analysis

Table 1 sets out some Monte Carlo experiments to investigate the properties of the estimators of β. We consider three values of β (.04, .02, .00) and four sets of initial conditions for the y_{it}, $-\lambda(1,2,...,20)$ for $\lambda = .5$, 1, 2, 4. A value of λ somewhere between 2 and 4 would characterise differences from US GDP (per head) in OECD data. Each of the 12 cells reports the results of 5000 experiments. The first entry in each cell gives the average of 5000 estimates of (3), the second entry the corresponding average for method (4), while the final entry gives the relative root mean squared error for the two methods. By construction, the v_{it} are orthogonal at different i and different t. In general both estimators are negligibly biased over the range of parameters considered but it is possible to identify several patterns.

Bias and estimate imprecision are increasing in β and decreasing in λ. For β=.04 the B-regression is always worse in Root Mean Square Error (RMSE) terms. For β=.02 or smaller this result is reversed except for small λ. Thus neither method is unambiguously superior. For λ=2, more or less the OECD value, ML is slightly better for β=.04 but becomes decidedly worse for smaller values. The reason for the relative diffuseness of the ML estimates is that they entail calculation of a large (20×20) covariance matrix of errors. If it is assumed in estimation that this covariance matrix is scalar (as is true in fact for the generated errors v_{it}), ML becomes better in terms of root mean square error, only by 10 - 14% for the values tabulated of λ and β.

However ML improves its performance when the v_{it} are correlated across i. Table 2 sets out some experiments varying the correlation (ψ) between the v_{it}. We find that ML dominates the B-regression once the average cross-correlation becomes greater than .25. Correlation of this order and greater characterizes many data sets, including per-capita GDP. We also include in Table 2 some experiment with a conventional pooling estimate of (4) (i.e. stacking and OLS, thus ignoring cross-sectional correlation). ML dominates pooling once cross-country correlation becomes greater than .25 .

3.1. Serial Correlation

We have noted above that the B–regression is more likely to be robust against the false imposition in estimation of parameter equality across conditioning variables. Robustness against mild mis-specification is a very useful property in this context. One likely form of mis-specification is the presence of serial correlation in v_{it} in (4). Indeed, Barro (1998) argues against panel estimates of (4) on the grounds that (4) pertains, not to GDP itself, but to GDP purged of its business cycle component. In this case, if observed GDP is used in estimation, bias is expected because of measurement error[13]. Specifically, assuming $y_{it} = \hat{y}_{it} + w_{it}$ where $\hat{y}_{it}$ is the value of per capita GDP appropriate to equation (4), the error term in (4) becomes

$$\varepsilon_{it} = v_{it} + w_{it} - (1-\beta) w_{it-1}$$

where w_{it} is the business cycle component of output. For real OECD data and taking values of β in the interval (0, 0.1) the average first-order serial correlation of the residual from (4) is positive and ranges between 0.2 and 0.3.[14] Naturally this would bias estimates of β from (4) towards zero. Barro (1998) and Shioji (1998) observe that measurement error related to the business cycle could bias β up. This is so but positive serial correlation in v is offsetting. As we have noted, for plausible values of β in OECD data, serial correlation in ε is positive, leading to bias in the opposite

---

[13] Note however that since y^*_{it} is a difference from a comparative country, common business cycle components are automatically eliminated.

[14] Thus, if v and w are independent AR(1)s, it must follow that the serial correlation in v_{it} is of the order of .2 to .3 on average.

direction. However to deal with measurement error Shioji (1998) proposes a "skipping" estimator[15]. The skipping procedure consists on estimating

$$y^{*}_{it} = (1-\beta)^{m} y^{*}_{it-m} + e_{mit} \tag{7}$$

by OLS, where $e_{mit} = v_{it} + (1-\hat{a})v_{it-1} + ... + (1-\hat{a})^{m-1} v_{it-m+1}$. A decreasing pattern on the estimated coefficient as m increases should suggest the presence of an important measurement error bias.

Another mis-specification that could have important effects occurs when β varies across countries or regions, a point made by Pesaran et al. (1997). If the variation is random then estimates of the average β derived from (3) will not be importantly biased[16]. However, as in the measurement error case, estimates of β from (4) will be biased towards 0 since parameter variability is more or less equivalent to adding an autocorrelated error. If however there is correlation in the sample between β_i and y^{*}_{io} then bias can be expected for both methods. The ML estimate of β derived from (4) may be thought of as a weighted average of OLS estimates of each i equation taken singly, wherein the weights are proportional to the sample variance of y^{*}_{it} (if the v_{it} are orthogonal). Thus if slow adjusters (low β) tend to be undeveloped regions (low y^{*}_{io}), as seems quite plausible, the estimate of β will be biased towards 0.

We study the effect of serial correlation on the β–coefficient by Monte Carlo in Table 3. The error processes are scaled to have unit unconditional variance (not unit innovation variance) and we denote the autoregressive parameter by α. Both estimators are relatively immune from bias, with the estimates derived from the B–regression somewhat more efficient. Note that, if the series y^{*}_{it} were realisations of stationary AR(1)s, we could expect strong biases in ML estimates of β in the presence of autocorrelated errors. The lack of bias in Table 3 is due completely to the initial conditions.

---

[15] See Shioji (1998) on this point. Using a skipping procedure, he finds evidence of possible measurement error in the US states and the Japanese prefectures. Evidence is less clearcut for the OECD.

[16] Though see Pesaran *et al.* (1997).

Table 4 gives the results of variability across i in the parameter β. We have taken 20 different values of β_i , evenly spread between .01 and .03. In terms of the usual economic interpretation, this is fairly extreme variability since at the lower end, gaps have a half-life of 70 years versus 23 years at the higher end. The first cell assumes the order of the βs corresponds to the initial conditions i.e. the lowest β is the least developed etc. Note that β is biased towards zero by about 0.005. The second cell reverses the assignment: the lowest βs go with the most developed. In this case the bias is reversed. It seems to us these results are quite reassuring since the tabulated biases are very much worst case outcomes.

Finally, Table 5 investigates the relationship between the initial conditions and the distribution of the ML estimator $\hat{\beta}$. We compute indices of skewness and kurtosis. We compute also confidence intervals for $\beta = .02$, defined as the points between the 2.5 percentile and the 97.5 percentile of these empirical distributions. It will be observed that $\hat{\beta}$ is effectively unbiased once λ rises above about 2.0 but the distribution still manifests skewness and excess kurtosis. Even for λ as high as 8.0, the distribution is leptokurtic. For low values of λ, 95% confidence intervals for λ are quite wide, indicating that tight estimates of convergence rates require wide variance in initial conditions.

3.2. Bias and Fixed Effects

A natural test in the general convergence proposition is to insert constants on the right in (4) i.e. to allow for fixed effects. If such constant are present in the true model, different countries will differ in per capita income in the long run, perhaps due to semi-permanent aspects of institutional and technological structure. In this case estimates of β from (4) will be biased towards zero if constants are not included. On the other hand, if constants *are* included, estimates of β will be biased up, whether the constants are present in the true model or not. In the univariate case, (6) bias is $(1+3\rho)/T$ when a constant is included. With K extra regressors on the right, the maximum bias is $[K(1+\rho) + 2\rho]/T$. In the multivariate case, allowing each country to have its own constant will bias β by approximately 4/40 = .10 given that the constants are in fact absent[17].

We analysed this latter bias by Monte Carlo. In the notation of Table 1 with β=.02, λ=2 and orthogonal v_{it}, we found with 5000 Monte Carlo replications an average estimate of .102 with sample standard deviation of .086 when each country was allowed its own constant. Examinations of parameter estimates for these artificial data showed that the constants were almost invariably statistically significant, despite being genuinely absent in the model. The magnitude of the fitted constants was correlated with starting conditions y^*_{io}, so that countries behind at t=0 were falsely predicted to be behind in equilibrium (t=∞). Reflection reveals why this must occur. Our initial conditions y^*_{i0} are negative and, since $\hat{\beta}$ is biased up, the predicted path (with a zero constant) for a given country from t = 0 to T must lie on average above the path in the data. It follows that the least-squares fit will be improved by choosing a negative constant. Thus bias to $\hat{\beta}$ creates bias to the fixed effects parameters.

This can also be seen heuristically by examination of (6). If $\hat{\rho}$ is derived with a fitted constant α then, since a regression line passes through the sample means, we must have

(8) $$z_t = \hat{\rho}\bar{z}_{t-1} + \hat{\alpha}$$

Thus, with no constant present in the true model, we have

(9) $$E(\hat{\alpha}) = E((\rho - \hat{\rho})\bar{z}_{t-1})$$

Given that $\rho - \hat{\rho}$ is almost always positive, it follows that, for countries whose GDP/head is negative over the whole sample, we expect to find negative constants.

---

[17] See Nickell (1981) and Islam (1995) for an application to the convergence equation.

4. Case Study: convergence in the OECD

As an example we have applied both methods of estimation to income per capita in 23 OECD countries (the whole set excluding Turkey and Yugoslavia) from the Penn world tables[18]. The sample thus consists of developed capitalist economies for whom the assumption of common tastes and technology is fairly supportable. The sample runs from 1950 to 1990. In 1950 these economies had been variously affected by the cataclysms of the first half of the century and thus are suitable for study of convergence in a homogeneous set of countries disturbed from equilibrium[19]. The United States was chosen as the comparator country.

Table 6 sets out estimates obtained by Shioji's skipping procedure. Equation (7) has been estimated using different values of m. The absence of any downward trend in the βs as m grows confirms the absence of measurement error bias in the estimates[20].

An OLS regression of $y^*_{iT} - y^*_{io}$ on y^*_{io} gave a parameter of -.39 with a standard error of .03 so the hypothesis of $\beta = 0$ in (4) is decisively rejected. The non-linear least squares estimate of β in (4) was $\beta = .023$ (.002). The estimate of β from (4) was quite different, $\hat{\beta} = .028$ (.0019). Which do we believe? The Durbin's h test for the i equations in (4) was .86 on average suggesting that bias arising from serial correlation in the residuals is not a problem[21]. In any case, this should bias the estimate towards zero. Similarly, variability across the β_s will bias the estimates towards zero if the undeveloped regions are slow adjusters, as seems the most likely case. It would seem therefore that ML is fairly free of the mis-specification biases we have considered. The sample correlation of the v_{it} was about 0.4 on average so, according to Table 2, ML on (4) is more efficient and to be preferred.

When we include individual constants on the right of (4) we found

---

[18] Version 5.6; Laspeyre chain linked, 1995, Table 10.1

[19] One might seek to argue that Portugal and perhaps Greece and Ireland were not developed economies in 1950, by such measures as proportion of the workforce in agriculture, literacy, and female education. Our results are not overly sensitive to this assumption.

[20] The inclusion of fixed effects in (7) does not affect this result.

[21] The average Durbin-Watson is 1.73, while the corresponding Durbin's h test gives a value of .86. Thus we can confortably accept the null of absence of first order serial correlation.

$\hat{\beta}$ = .079. The constants were always negative (the US always remains richer) and significant in all countries except Switzerland. The likelihood ratio test for the exclusion of all constants gave a reading of 77 which indicates a massive rejection ($\chi^2_{.01}$ = 40.3). However, the discussion in section 3.2 leads us to expect bias and falsely significant constants in a regression such as this. To illustrate this problem for the OECD, we generated artificial data taking y^*_{i0} at its 1950 value, with shocks uit down from a multivariate normal distribution with the covariance matrix of $u_{it} = y^*_{it} - 0.98y^*_{it-1}$ in the sample. The results are given in Table (7): the inclusion of constants, though genuinely absent in the true artificial data, introduces a substantial upward bias to $\hat{\beta}$.

Accordingly we have constructed a critical level for the likelihood ratio by Monte Carlo. We first estimated β under the null of no constants and computed the residual variance-covariance matrix. We then simulated (4) using random v_{it} with the computed covariance matrix and the initial conditions y^*_{io} observed in our data. Equation 3 was then estimated with and without constants. Repeating the process 5000 times, we found for the likelihood ratio test a 1% critical level of about 70 and a 5% level of 61.3. Thus unconditional convergence is comfortably rejected at conventional levels though nowhere near as decisively as with the conventional χ^2. The problem countries are Greece, Ireland, Portugal and Spain: their growth has been too low, given their relative poverty over the sample. Without these countries, absolute convergence is accepted at the 1% level by a conventional χ^2 test and at the 5% level by Monte Carlo. It is noteworthy that these four countries had the lowest secondary school enrolment in 1965 (World Development Report 1991, World Bank).

An alternative estimate of β in (4), robust to the presence of fixed effects can be based on the Anderson-Hsiao estimator wherein the equation is differenced and estimated using appropriate lags of y^*_{it} as instruments. We compute a standard error for this estimate taking into account the cross-correlation between country shocks[22]. We find

[22] Specifically, the asymptotic variance of the estimate is

$\hat{\beta}$ = .031 (.08). The point estimate is very close to that obtained by ML , despite being free of the possible bias introduced by missing fixed effects. Table (8) sets out the values for β obtained by the four methods we have considered.

5. Conclusion

One feature of the empirical literature on convergence is the abundance of different econometric techniques. Recent studies emphasise the advantages of panel estimates over the B–regression approach. So far, no one has considered cross-correlation of the error term. In this study we analyse by Monte Carlo the properties of the ML panel estimator with an unrestricted variance-covariance matrix in estimating the convergence parameter. We have found:

(i) both ML and the B–regression have various things in their favour. The B–regression is robust against some possible mis-specifications.

(ii) ML is a better estimator for cross-country correlation likely to be observed in most work.

(iii) the natural test for convergence in the ML approach of including country-specific constants has problems because bias to the lagged dependent variable, in conjunction with initial conditions, creates bias and false significance in the estimates of the constants.

(iv) Monte Carlo study of the likelihood ratio statistic suggests that the value obtained for the exclusion of the constants in the OECD data is statistically significant at conventional levels i.e. unconditional B–convergence is rejected in these data – but only just.

Table 1

Estimates of E(β) obtained from (3), from (4) & the root mean square errors of the estimates of (4) as a proportion of (3).

	β=*.04*	β=*.02*	β=*.00*

$\text{Var}(\hat{\beta}) = (\Delta y^{*}_{-1}{}'\Delta y^{*}_{-1})^{-2}\,(\Delta y_{-1}{}^{*}{}'(\Sigma\otimes J)\,\Delta y^{*}_{-1})$

where J the usual Anderson-Hsiao MA-matrix, and Σ is an estimate of the cross-sectional covariance matrix, given by the residual covariance matrix factored by 0.5. The vector Δy^{*}_{-1} is the stacked Δy^{*}_{-1}, instrumented by two lags of y^{*}_{-1}. For more on this point see Baltagi (1995).

λ=.5	.0415 (.0126)	.0212 (.0100)	.0009 (.0074)
	.1003 (.236)	.0252 (.0606)	.0007 (.0062)
	19.1	6.02	.83
λ=1	.0406 (.0074)	.0205 (.0058)	.0003 (.0042)
	.0407 (.0215)	.0200 (.0048)	.0002 (.0029)
	2.88	.82	.70
λ=2	.0401 (.0038)	.0201 (.0029)	.00010 (.0022)
	.0393 (.0040)	.0196 (.0023)	.00005 (.0014)
	1.06	.79	.68
λ=4	.0400 (.0019)	.0200 (.0015)	.00002 (.0010)
	.0391 (.0019)	.0195 (.0011)	.00003 (.0007)
	1.11	.82	.67

Notes:

(i) For each λ and each β the cells contain, respectively, the estimate of $E(\hat{\beta})$ from (4) (in the background), from (3) and the relative root mean square errors (*rmse*), (3)/(4) (in italics). Sample standard deviations are given in brackets. Equation (4) was estimated by maximum likelihood, with free covariance matrix of v_{it}. Equation (3) was estimated by non-linear least squares. Both equations were estimated using the LSQ option in TSP 4.2 (5000 replications).

(ii) Initial values of $(\Delta y_{1t},\ldots,\Delta y_{20}t) = \lambda(1,2,...20)$.

(iii) Convergence tolerance for estimator convergence = .0001. All calculations performed in double precision.

(iv) The covariance matrix of the v_{it} was taken to be I_{20} to generate the data.

Table 2
Estimates of bias and RMSE for different methods of estimating the convergence parameter β with cross correlations of the v_{it}
(λ=2, β=.02)

		ψ=.00	ψ=.25	ψ=.5	ψ=.75
(1)	ML	.0200 (.0029)	.0202 (.0043)	.0202 (.0040)	.0201 (.0030)
(2)	B-regression	.0195 (.0022)	.0202 (.0043)	.0205 (.0074)	.0209 (.0157)
(3)	Pooling	.0201 (.0020)	.0205 (.0051)	.0210 (.0060)	.0212 (.0072)
	rmse(2)/rmse(1)	*.77*	*1.16*	*1.86*	*5.27*
	rmse(3)/rmse(1)	*.68*	*1.01*	*1.52*	*2.46*

Notes:
(i) Convensions as for Table 1, (i)-(iii).
(ii) ψ is the cross-country correlation coefficient.
(iii) The Pooling estimator is OLS on the stacked data.

Table 3
Effect on E(β) of serial correlation in v_{it} (λ=2, β=.02)

	α=.25	α=.5
ML	.0196 (.0039)	.0183 (.0055)
B-regression	.0197 (.0030)	.0200 (.0046)
	.77	*.80*

Notes:
(i) Conventions as for Table 1, (i)-(iii).
(ii) α is the first order serial correlation coefficient

Table 4
Effect on *E*(β) of β–variability (λ=2)

high initial gap = slow adjusting	high initial gap = fast adjusting
.0155 (.0028)	.0249 (.0035)
.0150 (.0020)	.0245 (.0025)
.93	*.85*

Notes:
(i) Conventions as for Table 1, (i)-(iii).
(ii) The first cell computes β when β_i ranges from .03 to .01 for starting values -2(1,2, ..., 20); the second cell reverses this assignment.

Table 5
Properties of ML estimator as initial conditions vary (β=.02)

	mean	*Std. Dev.*	*skewness*	*Kurtosis (-3)*	*95% confidence intervals*
λ= .1	.0216	.0265	-.6280	1.1719	[-.0247, .0792]
λ= .5	.0221	.0160	-1.0061	3.2365	[-.0036, .0589]
λ= 1	.0209	.0083	-.6893	2.1411	[.0065, .0392]
λ= 2	.0202	.0042	-.4673	1.6036	[.0124, .0299]
λ= 4	.0200	.0021	-.2378	.9502	[.0160, .0243]
λ= 8	.0200	.0010	-.0695	.6628	[.0181, .0222]

Notes:
(i) Conventions as for Table 1, (i)-(iii).
(ii) Standard Errors for the skewness and kurtosis statistics are .04 and .08 respectively.

Table 6
Estimated β–parameter: results from skipping estimation

m	β *-coefficient*	*St. Dev*
1	.0265	(.0030)
2	.0275	(.0064)
4	.0280	(.0121)
8	.0263	(.0223)
10	.0277	(.0283)

Notes:
(i) Method of estimation: OLS (Pooling Estimation).
(ii) *m* defines different samples (data are taken every *m* years).

Table 7
Estimates of β with and without fixed effects in OECD data

	Mean	***Standard Deviation***	***Bias***
β (fixed effects fitted)	.079	.0262	.059
β (no fixed effects)	.020	.0049	.005

Notes:
(i) The Table gives the results of Monte Carlo estimations of b in artificial data, taking as initial conditions those observed in the OECD and covariance matrix corresponding to b=.02. There are no fixed effects in the artificial data.

Table 8
Sample OECD (1950-90)
Estimates of the convergence coefficient by different methods

	β *coefficient*	*Std. Error*
B-regression	.023	(.030)
OLS pooling	.027	(.003)
ML procedure	.023	(.002)
Fixed Effects	.068	(.005)
Anderson-Hsiao	.031	(.080)

Notes:
(i) Number of countries 22, number of observations 41. The sample includes the OECD countries with the exception of Yugoslavia and Turkey. Data are in difference from US levels.
(ii) The estimated value of the average Durbin's h test (ML procedure) is .86.

References

Baltagi B. H. (1995), *Econometric Analysis of Panel Data*, John Wiley and Sons , Chichester, UK.

Barro, R. J., Sala-i-Martin X. (1995), *Economic Growth*, McGraw Hill.

Baumol, W. J. (1986), Productivity Growth, Convergence and Welfare: What the Long-Run Data Show, *American Economic Review*, 76, 1072-85.

Caselli F., Esquivel G., Lefort F. (1996), Reopening the Convergence Debate: a New Look at Cross Country Growth Empirics, *Journal of Economic Growth*, 1, 363-89.

Durlauf S. N., Quah D. T. (1998), The Empirics of Economic Growth, *Centre for Economic Performance Discussion Paper* No. 384 (forthcoming in: (eds.), *Handbook of Macroeconomics*, North Holland).

Evans G.B.A., Savin N.E. (1983), Testing for Unit Roots: 2, *Econometrica*, 52, 1241-69.

Grubb D., Symons J. (1987), Bias in Regressions with a Lagged Dependent Variable, *Econometric Theory*, 3, 371-86.

Islam N. (1995), Growth Empirics: A Panel Data Approach, *The Quarterly Journal of Economics*, 110, 1127-70.

Knight M., Loayza N., Villaneuva D. (1993), Testing the Neoclassical Theory of Economic Growth, *IMF Staff Papers*, 40, 512-41.

Lee K., Pesaran H.M., Smith R. (1997), Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model, *Jounal of Applied Econometrics*, 12, 357-92.

Mankiw, G. N., Romer D., Weil N. (1992), A contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107, 407-37.

Nickell S. (1981), Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1399-416.

Peasaran H. M., Smith R. (1995), Estimating Long-run Relationships from Dynamics Heterogeneous Panels, *Journal of Econometrics*, 68, 79-113.

Robertson D., Symons J., (1992), Some Strange Properties of Panel Data Estimators, *Journal of Applied Econometrics*, 7, 175-89.

Shioji E. (1998), It's still 2%: Evidence on Convergence from 116 Years of the US States Panel Data, *Economic Working Paper* No.236, Universitat Pompeu Fabra, Barcelona.