

## GROWTH AND CONVERGENCE IN A MULTI-COUNTRY EMPIRICAL STOCHASTIC SOLOW MODEL

KEVIN LEE,<sup>a\*</sup> M. HASHEM PESARAN<sup>b</sup> AND RON SMITH<sup>c</sup>

<sup>a</sup>*Department of Economics, University of Leicester, University Road, Leicester LE1 7RH, UK  
E-mail: kcl2@leicester.ac.uk*

<sup>b</sup>*Trinity College, Cambridge, UK, and University of Southern California, Los Angeles, USA*

<sup>c</sup>*Birkbeck College, London, UK*

### SUMMARY

The paper considers international per capita output and its growth using a panel of data for 102 countries between 1960 and 1989. It sets out an explicitly stochastic Solow growth model and shows that this has quite different properties from the standard approach where the output equation is obtained by adding an error term to the linearized solution of a deterministic Solow model. It examines the econometric properties of estimates of beta convergence as traditionally defined in the literature and shows that all these estimates are subject to substantial biases. Our empirical estimates clearly reflect the nature and the magnitude of these biases as predicted by econometric theory. Steady state growth rates differ significantly across countries and once this heterogeneity is allowed for the estimates of beta are substantially higher than the consensus in the literature. But they are very imprecisely estimated and difficult to interpret. The paper also discusses the economic implications of these results for sigma convergence. © 1997 John Wiley & Sons, Ltd.

*J. Appl. Econ.*, **12**, 357–392 (1997)

No. of Figures: 2. No. of Tables: 7. No. of References: 26.

### 1. INTRODUCTION

The Solow–Swan neoclassical growth model, despite its age and recent developments in the growth literature, continues to be of great theoretical and empirical interest. There have been a large number of recent papers which test implications of this model, typically using data from the Penn World Tables (PWT) (Summers and Heston, 1991). Prominent examples are Mankiw, Romer, and Weil (1992), who use the data in cross-section, and Islam (1995), who uses the data in panel. Associated with this there has been considerable discussion of the meaning and interpretation of one of the central notions in the debate, namely ‘convergence’. The debate is discussed, for example, by the contributors to the ‘Controversy’ section of the July 1996 *Economic Journal* introduced by Durlauf (1996).

In this paper we use the results from an associated paper, Binder and Pesaran (1996) (BP) to develop an empirical version of a stochastic Solow growth model.<sup>1</sup> This differs significantly from the deterministic Solow model supplemented by an *ad hoc* stochastic specification which has been

---

\* Correspondence to: Kevin Lee, Department of Economics, University of Leicester, University Road, Leicester LE1 7RH, UK

Contract grant sponsor: ESRC; Contract grant number: H519 25 5003

Contract grant sponsor: Isaac Newton Trust of Trinity College, Cambridge.

<sup>1</sup> In their paper BP also consider the case where the saving rate is endogenously determined, as well as variants of ‘AK’ endogenous growth models.

characteristic of the previous literature. The stochastic model provides an alternative interpretation of the results previously obtained in the literature and suggests that some of the conclusions drawn from those results may be misleading. In addition to providing a tighter link between the specification of the economic theory and the empirical equation, we also use econometric theory to analyse the properties of the available estimators. The predictions of the economic and econometric theory for estimates of growth and convergence are then assessed on PWT data.

There are at least three notions of convergence in the literature. The first, beta convergence, considers the speed with which the logarithm of per-capita output, or output for short, tends to its steady-state value from some initial condition. The estimate of the speed of convergence is based on either the coefficient of the lagged dependent variable in time-series or panel regressions or the coefficient of the logarithm of initial output in cross-section regressions. Convergence may be unconditional or conditional on some country-specific variables, and could be to a common or to a country-specific steady-state growth rate. Thus the steady states for the logarithm of per-capita output may be identical (unconditional convergence with the same growth rate), parallel straight lines (conditional convergence with the same growth rate), or unrelated (conditional convergence with different growth rates). We shall show the importance of allowing for the heterogeneity of growth rates across countries in consistent estimation of the speed of convergence coefficient.

The second notion, sigma convergence, focuses on the behaviour of the cross-country variance of output over time. As has been noted by Friedman (1992) and Quah (1993), beta convergence is only a necessary and not a sufficient condition for output dispersion to reduce. Sigma convergence is theoretically interesting if one believes that there is a common equilibrium across countries, determined by shared global technologies and tastes, and that the speed of convergence to steady-state outputs is the same across countries. Otherwise, the movement of the cross-section variance of output over time will reflect initial conditions, the evolution of the dispersion of the country-specific equilibria and the rate of adjustment within each country. Each country could be converging to its own equilibrium, but cross-country equilibria could be diverging.

A third notion of convergence treats log per-capita output as an integrated variable and asks whether different countries share a common deterministic and/or stochastic trend. Bernard and Durlauf (1995, 1996) and Evans and Karras (1996) use this notion. We shall consider the role of such a common trend in our empirical work.

Section 2 sets out a stochastic version of the Solow growth model, drawing on BP where a detailed analysis of the properties of this stochastic growth model is provided. In the stochastic Solow model, the standard practice of linearizing around the deterministic steady state is not valid, and the traditional convergence coefficient no longer has its usual interpretation. The coefficient on which inference about convergence is normally based, namely that on lagged or initial output, will in general reflect a mixture of the convergence coefficient and the serial correlation coefficient of the technology shocks. Hence, a value of zero for the coefficient could arise either because of no beta convergence or a unit root in technology. Furthermore, the disturbances of this equation will contain a moving-average term, making unit root inference even more problematic than usual. In these circumstances, trying to determine whether there is a unit root against the alternative of a root very close to unity is likely to be impossible with available samples and techniques. Therefore, we conduct the analysis under both the assumptions of trend and difference stationarity.

Section 3 examines the econometric properties of the cross-section, pooled, and heterogeneous panel estimators of the speed of convergence when output is trend stationary but where allowance

is made for the possibility of heterogeneity of growth rates across countries. It shows that all these estimators are subject to important biases.

Section 4 presents empirical results under the various specifications suggested by the theoretical discussion. It shows that the estimates of the speed of convergence are very sensitive to the treatment of heterogeneity, particularly in the growth rate, and unrestricted estimates show much faster convergence than conventional estimates obtained, implicitly or explicitly, under the assumption of a homogeneous growth rate across countries. Despite this much faster estimated speed of convergence, we cannot reject the null of a unit root in output. We then discuss the implications of these results for sigma convergence.

The main results and conclusions of the paper are summarized in Section 5. These emphasize the importance of the heterogeneity in steady-state growth rates across countries (whether outputs have a unit root or not), both in its effect on the standard econometric approaches to estimating beta convergence and in its implications for sigma convergence. While the heterogeneity of growth rates is recognized in the literature, the results of this paper provide a formal statistical analysis to confirm its pervasiveness as well as establishing its importance for economic and econometric analysis of growth empirics.

The empirical analysis of the paper promotes the use of time-series methods applied to the individual countries' output series, treated as a heterogeneous panel. The application of time-series methods enables us to distinguish between the convergence of countries' outputs to their steady-state growth paths and the cross-sectional evolution of the paths themselves. Estimating cross-section regressions, or regressions using observations based on data averaged over long periods, makes it impossible to consider either the complex dynamic adjustments involved in the countries' output processes or the heterogeneity of growth rates across countries. By basing our empirical analysis on a stochastic version of the Solow model we are also able to provide a structural interpretation of the dynamics of the output process, and establish an explicit link between the stochastic processes of output and technology; in particular showing that the presence of a unit root in output does not necessarily contradict the stochastic version of the Solow model.

## 2. THE STOCHASTIC GROWTH MODEL

Consider a set of countries,  $i = 1, 2, \dots, N$ , over a number of years,  $t = 1, 2, \dots, T$ . Output in each country,  $Y_{it}$ , is produced by physical capital,  $K_{it}$ , and labour employed,  $L_{it}$ , through a Cobb–Douglas production function

$$Y_{it} = K_{it}^{\alpha} (A_{it} L_{it})^{(1-\alpha)} \quad 0 < \alpha < 1$$

$A_{it}$  represents technology and endowment while capital stock is given by

$$K_{it} = I_{i,t-1} + (1 - \delta)K_{i,t-1}$$

where  $\delta$  is the rate of depreciation. Investment,  $I_{it} = s_i Y_{it}$ , and the savings rate,  $s_i$ , is constant in each country. The evolution of capital per effective labour unit,  $k_{it} = K_{it}/A_{it}L_{it}$ , is then given by

$$\Delta \log k_{it} = -\Delta \log (A_{it} L_{it}) + \log (s_i k_{i,t-1}^{-(1-\alpha)} + 1 - \delta) \quad (1)$$

The stochastic processes determining technology and employment are

$$\log A_{it} = a_{0i} + g_i t + u_{ait} \quad (2)$$

$$u_{ait} = \rho_{ai} u_{ai,t-1} + \varepsilon_{ait} \quad |\rho_{ai}| \leq 1 \quad (3)$$

$$\log L_{it} = l_{0i} + n_i t + u_{bit} \quad (4)$$

and

$$u_{bit} = \rho_{bi} u_{bi,t-1} + \varepsilon_{bit} \quad |\rho_{bi}| \leq 1 \quad (5)$$

Both shocks to technology and labour inputs,  $u_{ait}$  and  $u_{bit}$ , allow for the possibility of a unit root. The technology shock summarizes all the factors that might shift total factor productivity, and the employment shock summarizes the outcome of the interaction between labour demand and supply influences. We have assumed a common production function parameter,  $\alpha$ , and depreciation rate,  $\delta$ , but all the other parameters, and in particular the initial endowment,  $a_{0i}$ , and the growth rate of technology,  $g_i$ , are allowed to differ across countries.<sup>2</sup> The parameter variations across  $i$  play a central role in the empirical analysis, but for clarity of exposition we shall first develop the model for a single country dropping the  $i$  subscript.

In deriving output equations to be used in the empirical analysis, we shall follow the standard procedure which linearizes a deterministic analogue of equation (1), by expanding it around the steady-state value of the effective capital–labour ratio and then substituting for capital and technology in terms of output to derive a univariate representation for output. The difference is that we shall apply this procedure to a stochastic rather than a deterministic growth model. Using equations (2) and (4) in equation (1) yields

$$\Delta \log k_t = -(n + g) - \Delta u_t + \log(sk_{t-1}^{-(1-\alpha)} + 1 - \delta) \quad (6)$$

where  $u_t = u_{at} + u_{bt}$  is a composite shock. We linearize equation (6) around  $E[\log(k_\infty)]$ , where  $k_\infty$  is the random variable that underlies the steady-state distribution of  $k_t$ . BP show that if the distribution of the shocks is appropriately truncated on the left so that large negative shocks are excluded, such a steady-state distribution exists and discuss the conditions under which it has moments, showing in particular that  $E[\log(k_\infty)]$  will depend on all the moments of  $u_t$ .

Since in steady state the expected value of  $\Delta \log k_t$  is zero, then taking expectations of equation (6) we also have

$$n + g = E[\log(sk_\infty^{-(1-\alpha)} + 1 - \delta)] \quad (6a)$$

Notice that the non-linear term in equation (6a) can be written as

$$\log\{se^{-(1-\alpha)\log(k_\infty)} + 1 - \delta\}$$

which is easily established to be a convex function of  $\log(k_\infty)$ . Therefore, by Jensen's inequality

$$\begin{aligned} n + g &= E(\log\{se^{-(1-\alpha)\log(k_\infty)} + 1 - \delta\}) \\ &= \log\{se^{-(1-\alpha)E[\log(k_\infty)]} + 1 - \delta\} + h \end{aligned} \quad (6b)$$

<sup>2</sup> Evans (1996a) notes that the share of capital differs substantially across countries, but the identification of  $\alpha$  with the share of capital may be problematic.

where  $h$  is a strictly positive number. The size of  $h$  depends on the degree of the curvature of the function in equation (6a) and the distribution of the shocks. Rewriting equation (6b) in terms of  $E[\log(k_\infty)]$ , we obtain:

$$E[\log(k_\infty)] = \frac{1}{1-\alpha} [\log(s) - \log(e^{n+g-h} - 1 + \delta)] \quad (7)$$

which can be used in the linearization of equation (6). Specifically, denoting the error of approximation by  $\xi_t$ , the expansion of the non-linear term in equation (6) around  $E[\log(k_\infty)]$  yields:

$$\log(sk_{t-1}^{-(1-\alpha)} + 1 - \delta) = \gamma - (1-\lambda)\log k_{t-1} + \xi_t \quad (8)$$

where

$$1 - \lambda = \frac{s(1-\alpha) e^{-(1-\alpha)E[\log(k_\infty)]}}{s e^{-(1-\alpha)E[\log(k_\infty)]} + 1 - \delta} > 0$$

and

$$\gamma = \log(se^{-(1-\alpha)E[\log(k_\infty)]} + 1 - \delta) + (1-\lambda)E[\log(k_\infty)]$$

Using equation (7),  $(1-\lambda)$  and  $\gamma$  simplify as follows:

$$1 - \lambda = (1-\alpha)[1 - (1-\delta) e^{-(n+g-h)}]$$

and

$$\gamma = n + g - h - [1 - (1-\delta) e^{-(n+g-h)}][\log(e^{n+g-h} - 1 + \delta) - \log(s)]$$

Also, for small values of  $n$ ,  $g$ ,  $\delta$ , and  $h$ ,

$$1 - \lambda \approx (1-\alpha)(n + g + \delta - h)$$

and

$$\gamma \approx (n + g - h) + (n + g + \delta + h)[\log s - \log(n + g + \delta - h)]$$

The linearization will only provide a reasonable approximation for small deviations from the mean of the distribution and it loses the interaction that exists between shocks and the trend which is an inherent feature of the non-linear model. The deterministic growth literature expands the non-linear expression around  $\log(k^*)$ , where  $k^*$  is the steady-state value of equation (6) with  $u_t = 0$ . In this case,  $(1-\lambda)$  reduces to the familiar measure of beta convergence in the deterministic case when  $h = 0$  (see, for example, Barro and Sala-i-Martin, 1995, p. 36). However, there is no reason to believe that, in general,  $h$  will be small, since the mean of the stochastic steady state will not correspond to the deterministic solution. Nor is there any reason to assume that the error of approximation,  $\xi_t$ , will be white noise. Typically, it will be serially correlated and its variance will go to zero as the steady-state distribution is approached.

Denote the logarithm of per-capita output,  $\log(Y_t/L_t)$ , by  $x_t$ , and  $\log(A_t)$  by  $a_t$ . Then the production function can be written as

$$x_t = a_t + \alpha \log k_t$$

and using equations (6) and (8), and assuming the error of approximation is relatively unimportant, we obtain the following relationship for the growth rate:

$$\Delta x_t = \Delta a_t + \alpha \left[ -(n + g) - \Delta u_t + \gamma - (1 - \lambda) \frac{x_{t-1} - a_{t-1}}{\alpha} \right]$$

which involves the unobservable technology variable,  $a_t$ . However,  $a_t$  can be eliminated using equations (2) and (3) to get:

$$x_t = \mu + (1 - \lambda)gt + \lambda x_{t-1} + e_t \quad (9)$$

where

$$e_t = (1 - \alpha)\Delta u_{at} - \alpha\Delta u_{bt} + (1 - \lambda)u_{a,t-1} \quad (9a)$$

and

$$\mu = \lambda g - \alpha h + (1 - \lambda) \left\{ a_0 + \frac{\alpha}{1 - \alpha} [\log s - \log(n + g + \delta - h)] \right\} \quad (9b)$$

which, apart from the structure of the composite error term,  $e_t$ , gives the standard result when  $h = 0$ . Equation (9a) makes it clear that technology and employment shocks have quite different impacts on the output process. In particular, even if there is a unit root in the time-series process for employment, this will not cause a unit root in the output process since the employment shock only appears in equation (9a) as a first difference. The same, however, is not true of the technological shock, and a unit root in the process generating technology shocks also causes a unit root in the output process.

To see this, using equations (3) and (5), we can substitute for  $\Delta u_{jt}$ ,  $j = a, b$ , in equation (9a) to obtain

$$e_t = -[(1 - \alpha)(1 - \rho_a) - (1 - \lambda)]u_{a,t-1} + \alpha(1 - \rho_b)u_{b,t-1} + (1 - \alpha)\varepsilon_{at} - \alpha\varepsilon_{bt}$$

In the context of the Solow growth model, where  $0 < \alpha < 1$  (and hence  $0 < \lambda < 1$ ), the output process will have a unit root only if  $e_t$  has a unit root. Suppose now that there is a unit root in the labour input process, but not in technology (i.e.  $\rho_b = 1$ ,  $|\rho_a| < 1$ ). Then  $e_t$  will be a linear function of stationary processes  $u_{a,t-1}$ ,  $\varepsilon_{at}$ , and  $\varepsilon_{bt}$ , and hence will be stationary itself. In contrast, when  $\rho_a = 1$ , then  $\Delta u_{at} = \varepsilon_{at}$ , and

$$e_t = (1 - \lambda)u_{a,t-1} + \alpha(1 - \rho_b)u_{b,t-1} + (1 - \alpha)\varepsilon_{at} - \alpha\varepsilon_{bt}$$

Hence,  $\{e_t\}$  will contain a unit root (through the unit root in  $u_{a,t-1}$ ), irrespective of whether  $\rho_b = 1$  or not. In what follows, to simplify the analysis, we assume  $\rho_b = 1$ , which seems reasonable in view of the time series evidence on employment, and write  $\rho$  for  $\rho_a$ . Equation (9) then becomes

$$x_t = \mu + (1 - \lambda)gt + \lambda x_{t-1} - [(1 - \alpha)(1 - \rho) - (1 - \lambda)]u_{a,t-1} + (1 - \alpha)\varepsilon_{at} - \alpha\varepsilon_{bt} \quad (10)$$

Substituting  $u_{at}$  and  $u_{a,t-1}$  from this equation in (3) we obtain:

$$x_t = [\mu(1 - \rho) + \rho(1 - \lambda)g] + (1 - \lambda)(1 - \rho)gt + (\lambda + \rho)x_{t-1} - \lambda\rho x_{t-2} \\ + (1 - \alpha)\varepsilon_{at} - (\lambda - \alpha)\varepsilon_{a,t-1} - \alpha\varepsilon_{bt} + \alpha\rho\varepsilon_{b,t-1} \quad (11a)$$

Under certain restrictions on the correlation between the employment and technology shocks, the error term in equation (11a) will have a first order moving-average structure which we represent by  $\varepsilon_t - \phi\varepsilon_{t-1}$ , where  $\varepsilon_t$  is a serially uncorrelated composite error term, and  $\phi$ , the moving-average coefficient, is a non-linear function of the variance–covariances of technology and employment shocks and the other parameters of the model. For small correlations between the two shocks, conventional estimates of the parameters (e.g.  $\alpha \approx 0.33$ ,  $n \approx 0.01$ ,  $g \approx 0.02$ ,  $\delta \approx 0.05$  and  $\rho \approx 0.9$ ) and a small value of  $h$ , the moving-average term is likely to be negative and close to minus 1.<sup>3</sup> Thus we can write the output equation with its MA component having a negative coefficient as

$$(1 - \lambda L)(1 - \rho L)x_t = [\mu(1 - \rho) + \rho(1 - \lambda)g] + (1 - \lambda)(1 - \rho)gt + (1 - \phi L)\varepsilon_t \quad (11b)$$

where  $L$  is the lag operator. This equation clearly shows that in a stochastic version of the Solow growth model, the linearized output process follows an ARMA(2,1). A third equivalent form is the augmented Dickey–Fuller representation:

$$\Delta x_t = [\mu(1 - \rho) + \rho(1 - \lambda)g] + (1 - \rho)(1 - \lambda)gt \\ - (1 - \lambda)(1 - \rho)x_{t-1} + \lambda\rho\Delta x_{t-1} + \varepsilon_t - \phi\varepsilon_{t-1} \quad (11c)$$

which establishes a direct link between the null hypothesis of a unit root in the output process and the underlying parameters of the stochastic Solow growth model.

This formulation has a number of important implications. First,  $1 - \lambda \approx (1 - \alpha)(g + n + \delta - h)$  is not the same as the traditional speed of convergence coefficient unless  $h = 0$ . The term  $h$  will depend on all the structural parameters of the model and the moments of  $u_t$ , which could vary over countries and perhaps over time. Second,  $\lambda$  is not separately identified from  $\rho$ , since they enter the autoregressive part of equation (11b) symmetrically. Third, the presence of both autoregressive and moving-average components in the output process means that there are likely to be common factors, thus leading to multiple maxima in the likelihood function. Fourth, using the ADF representation (11c), the null hypothesis that there exists a unit root in output is given by  $(1 - \rho)(1 - \lambda) = 0$ , or the product of three small numbers,  $(1 - \rho)$ ,  $(1 - \alpha)$  and  $(g + n + \delta - h)$ , being equal to zero. Thus, even when there is both convergence (i.e.  $0 < \lambda < 1$ ) and a stationary technology process ( $|\rho| < 1$ ), the coefficient of the lagged dependent variable in the ADF regression (11c) is likely to be indistinguishable from zero with the samples and techniques

<sup>3</sup> If the covariance between employment and technology shocks is zero, then equating  $E(\varepsilon_t)$  and  $E(\varepsilon_t\varepsilon_{t-1})$  with the corresponding expressions for the errors in equation (11a), provides the following:

$$\frac{\phi}{1 + \phi^2} = \frac{\left(\frac{\lambda - \alpha}{1 - \alpha}\right) + \rho\left(\frac{\alpha}{1 - \alpha}\right)^2 R}{1 + \left(\frac{\lambda - \alpha}{1 - \alpha}\right)^2 + [1 + \rho^2]\left(\frac{\alpha}{1 - \alpha}\right)^2 R}$$

where  $R$  is the ratio of the variance of technology shocks to that of employment shocks. Noting that  $(\lambda - \alpha)/(1 - \alpha) \approx 1 - (n + g + \delta - h)$ , then with conventional values of  $n$ ,  $g$ ,  $\delta$ ,  $\alpha$ , and  $\rho$ , and assuming  $h$  is relatively small we have  $\phi/(1 + \phi^2) \approx \frac{1}{2}$  and  $\phi \approx 1$ .

available. Fifth, it is clear from regression (11c) that a unit root in output is not evidence against the neoclassical model. The coefficient of  $x_{t-1}$ ,  $(1 - \rho)(1 - \lambda)$ , could be zero either because of lack of convergence, namely  $\alpha = \lambda = 1$ , or because of a unit root in technology,  $\rho = 1$ . Finally, the presence of the MA error in the output process complicates the problem of testing for unit roots since it is well known that unit root inference in the presence of MA errors with a negative root close to unity particularly lacks power.

In some circumstances, it may be possible to reduce the ARMA(2,1) process to an AR(1) process by eliminating a possible common factor from the model. If we can treat  $(1 - \lambda L)$  and  $(1 - \phi L)$  as approximately equal in equation (11b), and remove the common factor we get

$$x_t = d + (1 - \rho)gt + \rho x_{t-1} + \varepsilon_t \quad (12a)$$

where

$$d = \frac{\mu(1 - \rho)}{1 - \lambda} + \left( \rho - \frac{\lambda}{(1 - \lambda)^2} \right) g$$

which is the standard equation, except that the coefficient of the lagged dependent variable is interpreted not as the speed of convergence but the serial correlation coefficient in the technology process. Alternatively, technology may be highly correlated so that  $(1 - \rho L)$  is also of the same order of magnitude as  $(1 - \phi L)$  and we might remove this common factor to obtain

$$x_t = \mu + (1 - \lambda)gt + \lambda x_{t-1} + \varepsilon_t \quad (12b)$$

In this case, the coefficient of the lagged dependent variable provides a measure of the speed of convergence, and this is the interpretation typically adopted in the literature. In the empirical section below, we test for the presence of a common factor. When the common factor can be removed, we adopt the interpretation in equation (12b) for comparability with the literature. However, given the complexity of the univariate dynamics of output, we should emphasize that this is not the only possible interpretation.

For estimation, we need to reintroduce the distinction between countries,  $i = 1, 2, \dots, N$ . Suppose that we can remove a common factor from equation (11b) and interpret the model in terms of (12b), which we rewrite as

$$x_{it} = \mu_i + \theta_i t + \lambda_i x_{i,t-1} + \varepsilon_{it} \quad \begin{array}{l} i = 1, 2, \dots, N \\ t = 1, 2, \dots, T \end{array} \quad (13)$$

where  $\theta_i = (1 - \lambda_i)g_i$ , and from equation (9b),

$$\mu_i = \lambda_i g_i - \alpha h_i + (1 - \lambda_i) \left\{ a_{i0} + \frac{\alpha}{1 - \alpha} [\log s_i - \log(n_i + g_i + \delta - h_i)] \right\}$$

To separate the effects of  $\lambda_i$  and  $g_i$ , it is convenient to rewrite equation (13) as

$$x_{it} = c_i + g_i t + u_{it} \quad (14)$$

$$u_{it} = \lambda_i u_{i,t-1} + \varepsilon_{it} \quad (14a)$$

where  $c_i$ , the deterministic component of initial output, is defined by

$$c_i = a_{i0} + \frac{\alpha}{1 - \alpha} [\log s_i - \log(n_i + g_i + c - h_i)] - \frac{\alpha h_i}{1 - \lambda_i} \quad (14b)$$



The next section discusses estimation of these equations. For purposes of estimation also note that

$$\mu_i = (1 - \lambda_i)c_i + \lambda_i g_i \quad (15)$$

In discussing cross-sectional properties of the model we also use the notation  $x_{i0}^*$  for  $c_i$ , to highlight the fact that  $c_i$  is the deterministic component of the initial output,  $x_{i0}$ .

### 3. ECONOMETRIC CONSIDERATIONS

The  $N \times T$  system of equations in (13) is a simple example of the dynamic heterogeneous panels examined in Pesaran and Smith (1995) and Pesaran, Smith, and Im (1996). These papers discuss four ways such panels can be examined empirically: averaging the data over time and estimating a cross-section across groups; averaging the data over groups and estimating an aggregate time series; pooling the data assuming coefficient homogeneity; or estimating the coefficients of individual time series regressions and examining the distribution over groups of these estimates. We shall examine the econometric properties of these four approaches.

#### 3.1. Cross-section Data

The pure cross-section approach involves averaging the observations for each country over time and typically regressing the logarithm of per capita output on regressors such as  $\ln(s_i)$  and  $\ln(n_i + g + \delta)$ . Such cross-sectional regressions tell us nothing about the dynamic process of growth or convergence. To make inferences about convergence much of the literature estimates a hybrid cross-section, 'Barro', regression by including initial output to pick up dynamics. To examine the properties of this procedure assume a common  $\lambda$  ( $\lambda_i = \lambda$ ) and iterate equation (13) forwards from  $t = 1$  to  $t = T$  to obtain

$$x_{iT} - x_{i0} = -(1 - \lambda^T)x_{i0} + (1 - \lambda^T)x_{i0}^* + Tg_i + \xi_{iT} \quad (16)$$

where  $x_{i0}^*$  is the deterministic component of initial output, defined by (14b), and

$$\xi_{iT} = \sum_{j=0}^{T-1} \lambda^j \varepsilon_{i,T-j}$$

Analogues of equation (16) have been widely investigated using cross-country and regional data. This literature assumes a deterministic underlying growth model, sets  $h_i = 0$  and ignores the possible heterogeneity of  $g_i$ , the rate of growth of technology.<sup>4</sup>

While it is known that this procedure yields a biased estimate of  $\lambda$  (see, for example, Evans, 1996a), the number of different factors that could give rise to such a bias has not been recognized. Appendix A derives an explicit formula for the asymptotic (large  $N$ ) bias of the estimated speed of convergence based on equation (16), under the assumption that there is a common  $\lambda$ , but otherwise using very general assumptions. The expression for the asymptotic bias of estimating

<sup>4</sup> The assumption that  $h_i = 0$  is particularly important in cross-sectional analysis. In panel or time-series analysis the intercepts take account of the variations in  $h_i$  across  $i$ . For a cross-sectional analysis where the empirical implications of a non-zero  $h_i$  are investigated see BP.

$\lambda^T$  using Barro's regression (16) is derived in equation (A16), which is reproduced here for convenience:

$$\text{Plim}_{N \rightarrow \infty} (\widehat{\lambda^T}) - \lambda^T = \frac{(1 - \lambda^T)\sigma_{*0}^2 + T\sigma_{*g} + \sum_{j=1}^T \sum_{k=0}^{\infty} \lambda^{T-j+k} \bar{\gamma}(j+k)}{\sigma_{*0}^2 + \left(\frac{1}{1 - \lambda^2}\right) \left\{ \bar{\gamma}(0) + 2 \sum_{k=1}^{\infty} \lambda^k \bar{\gamma}(k) \right\}} \quad (17)$$

This shows that the magnitude of the bias depends on:  $\bar{\gamma}(k)$ , the  $k$ th-order autocovariance of the shocks averaged over  $i$ ;  $\sigma_{*0}^2$ , the cross-section variance of the deterministic component of initial output,  $x_{i0}^*$  defined by equation (14b), and  $\sigma_{*g}$ , the covariance between the deterministic component of initial output,  $x_{i0}^*$ , and the growth rate of technology,  $g_i$ .<sup>5</sup> This last term, which has previously been neglected, plays a crucial role since it is multiplied by  $T$ . The bias will either increase or decrease with the span of the regression, depending on whether  $\sigma_{*g}$  is positive or negative. It is highly unlikely for this covariance to be zero, since as can be seen from equation (14b),  $x_{i0}^*$  is itself a function of  $g_i$ .

Even in the very unlikely event that the growth rates of technology were the same across countries and therefore uncorrelated with the deterministic component of initial output, and  $\varepsilon_{it}$ 's were serially uncorrelated (so  $\bar{\gamma}(k) = 0$  for  $k > 0$  and  $\bar{\gamma}(0) = \tau^2$ ), Barro regressions will still yield biased estimates and the size of the bias will depend inversely on the ratio of the long-run time-series variance to the cross-section variance ( $\tau^2/(1 - \lambda^2)/\sigma_{*0}^2$ ). Empirical estimates indicate that this ratio is quite small even in 'conditional' Barro regressions, thus the bias is likely to remain serious even in this rather special case. Estimates of the magnitude of the bias based on panel regressions is discussed in Section 4.2.

### 3.2. Pooled Data

Under the assumption of a common growth rate ( $g_i = g$ ), a common rate of convergence ( $\lambda_i = \lambda$ ), and a common variance for  $\varepsilon_{it}$  across  $i$ , the appropriate way to estimate  $\lambda$  in equation (13) is to pool the data and use the traditional pooled fixed-effects estimation procedure. This approach has been followed in a recent paper by Islam (1995) who estimates an output equation using the Summers and Heston data, averaged over five five-yearly time spans.<sup>6</sup> For the fixed-effects estimates (provided in Table IV of his paper), Islam obtains implied estimates of  $\lambda$  of 0.9533 (0.0088), 0.9542 (0.0097), and 0.9074 (0.0157) for his Non-Oil, Intermediate and OECD country groupings, respectively.<sup>7</sup> These estimates imply faster rates of convergence than those obtained from Barro-type regressions, but convergence would still take place only after many years. Similar conclusions are obtained by Nerlove (1996) working in the context of a homogeneous dynamic random effects model. However, the validity of these estimates *critically* depends on the assumption of the presence of common growth rates across countries. The pooled fixed-effects estimator will yield inconsistent estimates of  $\lambda$  if the  $g_i$  differ across countries even if

<sup>5</sup> The derivation of equation (17) also allows for the possibility of a common (stochastic) component in the process generating the shocks,  $v_{it}$ . See equation (18).

<sup>6</sup> Other recent studies that also make use of panel data methods in their analysis of the growth process are Canova and Marcet (1995) and Miller (1995).

<sup>7</sup> Note that the implied estimates of the convergence rate reported in Table IV of Islam (1995) refer to the value of  $(1 - \lambda)$  in our notation.

$N$  and  $T$  are large. This is because imposing homogeneity of growth rates across countries when it is false adds a term,  $(g_i - g)t$ , to the disturbance for each country. This term is serially correlated with a unit coefficient, causing the dynamic fixed-effect estimator to be inconsistent for the reasons discussed in Pesaran and Smith (1995). The derivations in that paper assume that the regressors are stationary and are thus not appropriate to this case. However, Appendix B derives the appropriate asymptotic limits of the fixed-effect estimators of  $\lambda$  and  $\theta$ , denoted by  $\hat{\lambda}_{FE}$  and  $\hat{\theta}_{FE}$ , respectively, under heterogeneity of growth rates. These are:

$$\text{Plim}_{N,T \rightarrow \infty} (\hat{\lambda}_{FE}) = 1 \quad \text{Plim}_{N,T \rightarrow \infty} (\hat{\theta}_{FE}) = 0$$

irrespective of their true values, or whether  $N$  or  $T$  is allowed to go to infinity first.

In small samples, this upward heterogeneity bias will be combined with the familiar downward-lagged dependent variable bias. Under slope homogeneity, the fixed-effect estimator is inconsistent for  $N$  large and fixed  $T$ . There are a variety of instrumental variable estimators available in the literature which are consistent for fixed  $T$  and large  $N$  under slope homogeneity (a review is provided in Baltagi (1995, Chapter 8)). However, in general these estimators are not consistent under slope heterogeneity, as is discussed in Pesaran, Smith, and Im (1996). In fact, when slope coefficients differ across groups, there exists no consistent instrumental variables estimator (for a proof, see Pesaran and Smith, 1995).

### 3.3. Heterogeneous Panel Data

Under slope heterogeneity, one appropriate estimator is the mean group estimator. The equation is estimated separately for each country and the distribution of the coefficients examined. Evans (1996a) uses this procedure allowing for heterogeneity in  $\lambda_i$ , but not in  $g_i$ . However, the small-sample downward bias in the coefficient of the lagged dependent variable remains a serious problem. Since each of the country estimates are subject to this bias, it will not be reduced by averaging across countries or pooling. Evans tries to deal with this problem by using a median unbiased estimator, but such estimators may have a large variance.

Another important consideration in the analysis of dynamic heterogeneous panels is the possibility that the errors in equations (13) or (14) are correlated across countries. Such cross-sectional error correlations must be taken into account in empirical analysis. A simple and effective procedure would be to demean the observations before estimation. The demeaning procedure can be justified in the context of equation (14) if  $\lambda_i = \lambda$ , and the cross-sectional dependence of  $\varepsilon_{it}$  across  $i$  can be specified by the two-factor model:

$$\varepsilon_{it} = \eta_t + v_{it} \quad (18)$$

where  $\eta_t$  is a time-varying common (stochastic) component, and  $v_{it}$  is the country-specific disturbance term assumed to be independently distributed across  $i$ . For this specification, the demeaned version of equation (13) is given by

$$x_{it} - \bar{x}_t = (\mu_i - \bar{\mu}) + (\theta_i - \bar{\theta})t + \lambda(x_{i,t-1} - \bar{x}_{t-1}) + v_{it} - \bar{v}_t \quad (19)$$

where

$$\bar{x}_t = N^{-1} \sum_{i=1}^N x_{it}, \quad \bar{\mu} = N^{-1} \sum_{i=1}^N \mu_i$$

etc. It is now easily seen that for large enough  $N$ , the demeaned disturbances  $\varepsilon_{it} - \bar{\varepsilon}_t = v_{it} - \bar{v}_t$  are uncorrelated across  $i$ . In the case where  $\varepsilon_{it}$ 's are normally distributed, the errors in the demeaned regression will also be independently distributed across  $i$ , for large  $N$ .

Analogous to equation (14), it is also possible to rewrite equation (19) as

$$x_{it} - \bar{x}_t = (c_i - \bar{c}) + (g_i - \bar{g})t + u_{it} - \bar{u}_t \quad (20)$$

where

$$u_{it} - \bar{u}_t = \lambda(u_{i,t-1} - \bar{u}_{t-1}) + v_{it} - \bar{v}_t \quad (20a)$$

and  $c_i$  is defined as before by equation (14b). Finally, note that estimation of the demeaned regressions (19) or (20) yield estimates of  $c_i - \bar{c}$ ,  $g_i - \bar{g}$ , etc. To obtain estimates of  $c_i$ ,  $g_i$ , and  $\mu_i$  we also need to estimate the aggregate equation

$$\bar{x}_t = \bar{c} + \bar{g}t + \bar{u}_t \quad (21)$$

where

$$\bar{u}_t = \lambda\bar{u}_{t-1} + \bar{\varepsilon}_t \quad (21a)$$

Estimates of  $c_i$  and  $g_i$  can now be recovered from those of  $c_i - \bar{c}$  and  $g_i - \bar{g}$  in equation (20), and  $\bar{\delta}$  and  $\bar{g}$  in (21). The estimate of  $\mu_i$  can then be obtained using equation (15).

The above derivations clearly show that the demeaning procedure works exactly when equation (18) holds and  $\lambda_i = \lambda$ . However, even if  $\lambda_i$ 's differ across countries, but not markedly so, demeaning before estimation can still help remove some of the correlations that may exist across  $\varepsilon_{it}$ 's. Islam (1995) and Evans and Karras (1996) also demean the data in this way, but motivate it as removing a common, possibly stochastic, technology trend. Neither allow for differing country-specific growth rates as do equations (19) or (20).

This brief econometric review suggests that there are very real difficulties in obtaining precise estimates of the speed of convergence. If the Solow model is true with a common growth rate of technology across countries, the estimate of  $\lambda$  from Barro cross-section-type regressions will be biased upwards, suggesting slower convergence than is the case, and the appropriate estimator to be used is the pooled fixed-effect estimator. But if growth rates differ, the estimates of  $\lambda$  from the pooled fixed-effect procedure will tend to unity as  $N$  and  $T$  grow, irrespective of the true value of  $\lambda$ . When the growth rates differ, the mean group estimator is appropriate, but the estimate of the mean  $\lambda_i$  across  $i$  from this procedure is subject to a downward bias in small samples and this bias can be important even for  $T$  as large as 30.

#### 4. EMPIRICAL ESTIMATES OF THE SOLOW MODEL

This section presents estimates of the theoretical models of the univariate processes for the logarithm of per-capital output across countries described in Section 2. The parameters of interest are the country-specific technology growth rates,  $g_i$ , and the speed of adjustment to steady state,  $1 - \lambda_i$ . Different country-specific intercepts are always included which will capture such factors as differences in steady-state growth rates, rates of growth of population, initial endowments, differences in  $h_i$ , etc. We shall conduct inferences using three different specifications. The most general is the ARMA(2,1) with trend, given in equation (11b). However, the

theory suggests that the MA coefficient is likely to be close to *minus 1*, and that a common factor could be removed. Following this route gives the second specification, the stationary AR(1) model with trend given by equation (13) or, equivalently, by (14). However, the theory suggests that it may be difficult to reject the null of a unit root in this specification. Therefore this hypothesis will be tested and the third specification will impose the unit root restriction in order to examine the robustness of our main conclusion concerning growth rate heterogeneity to the unit root specification.

#### 4.1. Data

In testing the theory, like many others, we will use the Penn World Tables (PWT), described in Summers and Heston (1991).<sup>8</sup> We shall examine 102 non-oil-producing countries and two subsets of them (an intermediate group of 61 countries which excludes countries which are small or for which the quality of the data are thought to be poor, and a group of 22 OECD countries<sup>9</sup>). For these countries there are data from 1960 to 1989. The dataset is constructed on the basis of information both from national accounts and from a set of benchmark United Nations International Comparison Programmes. The focus of the analysis is on the logarithm of real per capita GDP in these country groupings over the period 1960–89, and this is measured by national accounts data which is scaled to ensure comparability across countries in 1985. The output equations are then estimated over the period 1965–89 (inclusive) to ensure comparability across different dynamic specifications.

As Heston and Summers (1996) emphasize, there are many deficiencies in this data, particularly for the poorer countries and those that have never been benchmarked. In some cases, the annual data which we use are interpolated. However, the fact that very similar features emerge from all three samples, with very different degrees of measurement problem, suggests that our main conclusion, the sensitivity of convergence estimates to the treatment of growth rate heterogeneity, is robust to such deficiencies.

#### 4.2. The Trend Stationary Case

The most general model we consider is the ARMA(2,1) process with a linear trend. This was estimated for the 102 countries and the theory's prediction of an MA root of  $-1$  was confirmed in 76 of the countries, with no obvious pattern across subsamples.<sup>10</sup> Of the remaining 26 countries, six have a root at  $+1$ , while of the 20 countries with an interior maximum, the common factor restriction was rejected in eight cases. Therefore, only eight out of the 102 countries seem to require an ARMA(2,1) specification. These countries were Rwanda, Costa Rica, El Salvador, Peru, the Phillipines, Iceland, Norway, and Sweden. Given these results, we decided to work with the more parsimonious AR(1) model containing a linear trend, and thus eliminated the common factor that seems to exist in the ARMA(2,1) specifications.

Four variants of the AR(1) model were estimated. Table I gives the unrestricted model where both  $g_i$  and  $\lambda_i$  are allowed to differ across countries. Table II gives the results when  $g_i$  is

<sup>8</sup> Version 5.0 of the PWT is described in Summers and Heston (1991). The PWT dataset that we use is Version 5.6, dated 20 November 1994, and the measure of per-capita output we use is labelled RGDP.

<sup>9</sup> It is worth noting that the OECD is a club which countries join only when they become rich. Therefore, analysis of the growth in per capita output in this group may be subject to considerable sample selection bias, and should be treated with caution.

<sup>10</sup> For example, 15 of the 22 OECD countries had an MA root of  $-1$ .

Table I. Summary statistics of estimated coefficients from the Solow growth model (calculated using demeaned data)

	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{g}_i$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
<i>(LL = 4495.84)</i>				
Mean	0.3255	0.7013	0.01742	0.04985
(Standard error)	(0.0335)	(0.0183)	(0.0020)	(0.0028)
Median	0.3485	0.7462	0.01924	0.03990
Standard deviation	0.3370	0.1825	0.02008	0.02775
Minimum	-1.0415	0.0412	-0.03106	0.01305
Maximum	1.0283	0.9253	0.07411	0.15945
<i>Intermediate group of countries (N = 61)</i>				
<i>(LL = 3078.15)</i>				
Mean	0.4078	0.7370	0.02243	0.03757
(Standard error)	(0.0344)	(0.0229)	(0.0022)	(0.0024)
Median	0.4050	0.7877	0.02242	0.03514
Standard deviation	0.2667	0.1773	0.01698	0.01865
Minimum	-0.7972	0.1549	-0.02085	0.01165
Maximum	1.0781	0.9254	0.07419	0.09923
<i>OECD countries (N = 22)</i>				
<i>(LL = 1404.98)</i>				
Mean	0.6806	0.7619	0.02652	0.02075
(Standard error)	(0.0246)	(0.0245)	(0.0017)	(0.0015)
Median	0.7038	0.7891	0.02623	0.02017
Standard deviation	0.1153	0.1121	0.00782	0.00691
Minimum	0.3046	0.5452	0.01238	0.00978
Maximum	0.8076	0.9142	0.04548	0.03612

*Notes:*

The estimates in this table are obtained by first estimating the demeaned regression equations (see equations (20) and (20a) in the text):

$$x_{it} - \bar{x}_i = c_i^d + g_i^d t + u_{it} - \bar{u}_i \quad \text{where } u_{it} - \bar{u}_i = \lambda_i(u_{i,t-1} - \bar{u}_{i,t-1}) + v_{it} - \bar{v}_i$$

separately for each country ( $i = 1, \dots, N$ ), over the period  $t = 1965, \dots, 1989$  by the exact ML method, where  $x_{it}$  is the logarithm of the per capita output in country  $i$ . Noting that  $c_i^d = c_i - \bar{c}$ ,  $g_i^d = g_i - \bar{g}$ , the estimates  $\hat{c}_i^d$  and  $\hat{g}_i^d$  were then used in conjunction with the ML estimates of  $\bar{c}$  and  $\bar{g}$  (denoted by  $\hat{c}$  and  $\hat{g}$ ) in equation (21) to compute the estimates of  $\hat{c}_i = \hat{c}_i^d + \hat{c}$  and  $\hat{g}_i = \hat{g}_i^d + \hat{g}$ . Finally, estimates of  $\mu_i$  were obtained using equation (15), and  $\tau_i$  are estimated as standard errors of the individual country demeaned output equations. Computation of the ML estimates were carried out using the 'inverse interpolation' method (see Pesaran and Slater, 1980). *LL* is the maximized value of the log-likelihood function. The 'mean', 'median', 'standard deviation', 'minimum', and 'maximum' statistics refer to the cross-country distribution of the estimated parameters. The standard error of the mean values of the coefficients estimated across countries is based on the country-specific parameter estimates and is calculated using the non-parametric procedure described in Pesaran, Smith, and Im (1996).

constrained to be the same across countries, Table III when  $\lambda_i$  is the same, and Table IV when both  $g_i$  and  $\lambda_i$  are the same. Estimation is carried out by the exact maximum likelihood method, constraining  $\lambda_i$  to be less than unity, and allowing the variances to differ across countries. Appendix C describes the procedure in detail. Ordinary Least Squares (OLS) estimates were broadly similar except that the OLS estimates of  $\lambda_i$  were sometimes greater than unity. The estimates in Table IV are subject to the same slope coefficient restrictions as the traditional fixed-effects panel estimator, but allow for the possibility of heterogeneity in variances of the shocks

Table II. Summary statistics of estimated coefficients from the Solow growth model under  $H_0: g_i = g$  (calculated using demeaned data)

	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{g}_i$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
<i>(LL = 4077.33)</i>				
Mean	0.3727	0.8993	0.02389	0.05594
(Standard error)	(0.0134)	(0.0104)		(0.0030)
Median	0.3771	0.9357		0.04863
Standard deviation	0.1387	0.1045		0.03030
Minimum	-0.2575	0.4487		0.01302
Maximum	0.8296	0.9937		0.17452
<i>Intermediate group of countries (N = 61)</i>				
<i>(LL = 2850.26)</i>				
Mean	0.4039	0.8979	0.02466	0.04196
(Standard error)	(0.0167)	(0.0163)		(0.0027)
Median	0.4070	0.9418		0.03804
Standard deviation	0.1296	0.1263		0.02100
Minimum	-0.0740	0.3178		0.01152
Maximum	0.8328	0.9938		0.10912
<i>OECD countries (N = 22)</i>				
<i>(LL = 1339.56)</i>				
Mean	0.6716	0.9018	0.02540	0.02255
(Standard error)	(0.0180)	(0.0198)		(0.0017)
Mean	0.6886	0.9343		0.02113
Standard deviation	0.0844	0.0928		0.00796
Minimum	0.3090	0.6869		0.00957
Maximum	0.7394	0.9904		0.04066

*Notes:*

Estimated coefficients relate to regression equations of the form given in the notes to Table I, but estimated by the exact ML method subject to the restriction  $g_i = g$ , for all  $i$ . See notes to Table I for further details.

across countries. The equations were estimated in the form of equation (20), using demeaned data. Log-likelihood ratio tests indicated that demeaning always led to a significant improvement in fit.

Table I indicates that the average growth rate increases from 1.7% in the full sample, to 2.2% in the intermediate and 2.6% in the OECD sample. The standard deviation of the growth rate falls across samples, as does the average standard error of the equation, which is just under 5% in the full sample and just over 2% in the OECD. In the light of the consensus in the literature that the speed of convergence is very low (e.g. Barro and Sala-i-Martin, 1995, suggest it is around 2%), the most striking feature of the table is the estimate of the average value of  $\lambda_i$ . The estimate of 0.70 in the full sample implies a speed of convergence of 30%, although this is convergence to a country-specific steady state. The estimates for the intermediate group of countries, at 26%, and for the OECD countries, at 24%, are also much higher than the consensus estimates. Using the original, rather than demeaned, data gave slightly slower speeds of convergence with average values of the estimated  $\lambda_i$ 's rising from 0.76 in the full sample to 0.81 in the OECD.

The heterogeneity of growth rates is very evident from the descriptive statistics in Table I. In the full sample average steady-state growth rates over this period range from minus 3% to positive 7%. Comparisons of results in Tables I and II indicate that the restriction of homogeneous

Table III. Summary statistics of estimated coefficients from the Solow growth model under  $H_0: \lambda_i = \lambda$  (calculated using demeaned data)

	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{g}_i$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
<i>(LL = 4425.90)</i>				
Mean	0.3765	0.7746	0.01741	0.04799
(Standard error)	(0.0209)		(0.0020)	(0.0026)
Median	0.3545		0.01916	0.04014
Standard deviation	0.2097		0.02003	0.02670
Minimum	-0.0054		-0.02877	0.01233
Maximum	0.7791		0.07371	0.15476
<i>Intermediate group of countries (N = 61)</i>				
<i>(LL = 3039.34)</i>				
Mean	0.4093	0.8096	0.02241	0.03603
(Standard error)	(0.0207)		(0.0022)	(0.0023)
Median	0.3922		0.02218	0.03361
Standard deviation	0.1601		0.01716	0.01781
Minimum	0.1283		-0.02151	0.01200
Maximum	0.6815		0.07391	0.09310
<i>OECD countries (N = 22)</i>				
<i>(LL = 1396.67)</i>				
Mean	0.6815	0.7999	0.02643	0.01977
(Standard error)	(0.0190)		(0.0017)	(0.0014)
Median	0.7072		0.02624	0.01919
Standard deviation	0.0890		0.00771	0.00660
Minimum	0.4412		0.01212	0.00916
Maximum	0.7975		0.04439	0.03409

*Notes:*

Estimated coefficients relate to regression equations of the form given in the notes to Table I, but estimated by the exact ML method subject to the restriction  $\lambda_i = \lambda$ , for all  $i$ . See notes to Table I for further details.

growth rates is strongly rejected, with log-likelihood ratio (LR) statistics of 837.02, 455.78, and 130.84 in the three samples. These are distributed as Chi-squared variates with 101 (124.3), 60 (71.08), and 21 (32.67) degrees of freedom, respectively. The 95% critical values are in brackets. Comparison of Tables I and III shows that the evidence against homogeneity in speeds of adjustment is less strong, with LR statistics of 139.88, 77.62, and 16.62 with the same degrees of freedom. Thus homogeneity in speeds of adjustment is not rejected in the case of the OECD sample.

As the theory in Appendix B indicates, imposing homogeneity pushes the estimate of  $\lambda$  towards unity, and substantially reduces the measured speed of convergence. In the empirical applications imposing equality of growth rates raises the average estimate of  $\lambda_i$  to 0.90 in the full sample. With both common  $g$  and  $\lambda$ , the estimates of  $\lambda$  increases to 0.96, which is comparable to those obtained in panel studies that have imposed homogeneity of technology growth rates (as in Islam, 1995, and Nerlove, 1996). This suggests that the low estimates of the speed of convergence obtained elsewhere are a direct result of the heterogeneity bias.

As Appendix A shows, the cross-section estimates are also biased. It is instructive to use the formula for the bias in equation (17) and the empirical estimates in Table III, which assumes a



Table IV. Summary statistics of estimated coefficients from the Solow growth model under  $H_0: g_i = g, \lambda_i = \lambda$  (calculated using demeaned data)

	$\hat{\mu}_i$	$\hat{\lambda}_i$	$\hat{g}_i$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
<i>(LL = 3990.91)</i>				
Mean	0.3765	0.9628	0.01741	0.05448
(Standard error)	(0.0039)			(0.0028)
Median	0.3737			0.04690
Standard deviation	0.0371			0.02859
Minimum	0.3137			0.01599
Maximum	0.4440			0.16871
<i>Intermediate group of countries (N = 61)</i>				
<i>(LL = 2805.97)</i>				
Mean	0.4093	0.9614	0.02241	0.04109
(Standard error)	(0.0042)			(0.0026)
Median	0.4026			0.03722
Standard deviation	0.0329			0.02051
Minimum	0.3499			0.01286
Maximum	0.4635			0.10490
<i>OECD countries (N = 22)</i>				
<i>(LL = 1328.07)</i>				
Mean	0.6815	0.9477	0.02643	0.02201
(Standard error)	(0.0047)			(0.0017)
Median	0.6884			0.02061
Standard deviation	0.0217			0.00789
Minimum	0.6178			0.00940
Maximum	0.7085			0.03900

*Notes:*

Estimated coefficients relate to regression equations of the form given in the notes to Table I, but estimated by the exact ML method subject to the restrictions  $g_i = g$  and  $\lambda_i = \lambda$ , for all  $i$ . See notes to Table I for further details.

common  $\lambda$ , to provide an indication of the quantitative importance of the bias in the cross-section estimates of  $\gamma$  based on Barro regressions. Assume that there is no serial correlation in the demeaned disturbances,  $v_{it}$ , but allow growth rates to differ across countries so that they could be correlated with the deterministic component of initial output,  $x_{i0}^*$ . Suppose that the true value of  $\lambda$  is 0.77 and the determinants of the bias match their values for the 102-country PWT sample as derived from the estimates in Table III.<sup>11</sup> Then equation (17) predicts that as  $N \rightarrow \infty$ , the value of  $\lambda$  estimated from cross-section Barro regressions would be greater than unity, and increase with  $T$ . This is consistent with the general failure to find evidence of unconditional beta convergence in the non-oil PWT sample. Similarly, suppose that the true value of  $\lambda$  is 0.80 and the determinants of the bias match their values for the OECD sample derived from Table III;<sup>12</sup> then equation (17) predicts that the  $\lambda$  estimated from the cross-section Barro regression with

<sup>11</sup> In using equation (17) note that  $\bar{\gamma}(0) = \tau^2$ . The relevant estimates for the full 102 sample are:  $\hat{\sigma}_{*0}^2 = 0.84$ ,  $\hat{\tau}^2 = 0.0023$ , and  $\hat{\sigma}_{*g} = 0.00299$ .

<sup>12</sup> The relevant estimates in this case are  $\hat{\sigma}_{*0}^2 = 0.21$ ,  $\hat{\tau}^2 = 0.0004$ , and  $\hat{\sigma}_{*g} = -0.00223$ .

$T = 10$  would be 0.988, and should decline slowly with  $T$ , since  $\hat{\sigma}_{*g}$  is negative.<sup>13</sup> Thus there is a clear indication that even if there was rapid convergence in individual economies from some initial output levels, as our estimates suggest, traditional cross-section and panel procedures would not reveal it.

#### 4.3. The Difference Stationary Case

So far we have assumed that the output series can be adequately approximated by a trend stationary model and have obtained estimates of the speed of convergence of around 30%. Although the mean of the  $\hat{\lambda}_i$  are quite precisely estimated, it is nevertheless subject to the familiar downward lagged dependent variable bias. There is also a sizeable body of literature which shows that the hypothesis of a unit root in output cannot be rejected. This poses an important problem of how to reconcile the empirical evidence of a unit root in output with the relatively large estimates obtained for the mean of  $\lambda_i$ . This section addresses this issue, and discusses the robustness of our results concerning the heterogeneity of growth rates to the trend stationary assumption of the previous sections.

Turning to the ARMA(2,1) specification given by equation (11c) it is clear that the hypothesis of a unit root in output is equivalent to testing the null hypothesis that either  $\lambda = 1$  or  $\rho = 1$ . Immediately, this presents us with the identification issue already alluded to in Section 2. It is indeed possible for  $\lambda$  to be around 0.7, and yet leading to non-rejection of a unit root in the output process if  $\rho$  is close enough to unity. This seems in fact to be the case in the PWT sample that we consider. Using demeaned data and choosing the order of augmentation,  $p_i$ , of the output process by the Schwarz Bayesian Criterion (SBC), the ADF( $p_i$ ) statistics (computed for the trended case) rejected the unit root null in 14 of the 102 countries. Setting  $p_i = 4$  for all countries, the null was rejected in only six countries.

It is known that ADF tests have low power and Im, Pesaran and Shin (1995) (IPS) propose a test that increases power by exploiting the panel structure of the data. The standardized  $t$ -bar test they propose is based on the average value of the Augmented Dickey–Fuller statistic,  $t_{iT}(p_i, \hat{\gamma}_i)$ ,  $i = 1, 2, \dots, N$ , based on the ADF with trend regression of order  $p_i$ , on a sample of  $T$  observations, where  $\hat{\gamma}_i$  is the estimated vector of coefficients on the augmented lagged changes. The standardized  $t$ -bar statistic is then simply calculated as

$$\bar{z}_{NT} = \frac{\frac{1}{N} \sum_{i=1}^N t_{iT}(p_i, \hat{\gamma}_i) - \frac{1}{N} \sum_{i=1}^N E[t_T(p_i, 0)]}{\sqrt{\frac{1}{N^2} \sum_{i=1}^N V[t_T(p_i, 0)]}} \quad (22)$$

IPS show that under the null hypothesis, when  $N$  and  $T$  are large and  $\sqrt{N}/T$  small, this statistic has a standard normal distribution. Asymptotically, the effect of assuming that  $\gamma_i = 0$  in the computation of the expected values and the variances of the individual ADF statistics on the  $t$ -bar statistic will become negligible. The values of  $E[t_T(p_i, 0)]$  and  $V[t_T(p_i, 0)]$  are tabulated in IPS.

<sup>13</sup> Notice, however, that in the case of the OECD sample, the large  $N$  assumption required to derive equation (17) may not be valid.

Using the SBC criterion to determine  $p_i$ , the test statistics on the demeaned data using equation (22) were  $-2.628$  for the full sample, which rejects the null at standard significance levels,  $-0.442$  for the intermediate, and  $0.213$  for the OECD. Setting  $p_i = 4$  for all countries the test statistics were  $-1.146$ ,  $1.871$ , and  $0.983$ , none of which reject the null. Given the sensitivity of the tests to underestimation of the degree of augmentation the conclusion is that these test statistics provide little evidence to reject the unit root null.

Kwiatkowski *et al.* (1992) propose a test in which the null is stationarity and involves testing whether the variance of the stochastic trend component of the series is zero. The test is semi-parametric and requires choice of a truncation parameter, which sets the window size for the spectral density estimator that underlies the computation of the test statistics. The number of countries rejecting the null of stationarity fell from 96 when the truncation remainder was set at zero, to 51 when it was set at four, and to nine when set at eight. Thus stationarity is still rejected in the case of half of the countries in the sample unless the truncation parameter is set very high.

Tables V and VI provide the results of the estimates of the AR processes for demeaned output growth rates for the countries in our dataset, assuming that there is a unit root in the log per capita output series. The model is given by

$$\Delta x_{it} = g_i + u_{it} \text{ where } u_{it} = \gamma_{1i}u_{i,t-1} + \gamma_{2i}u_{i,t-2} + \varepsilon_{it} \quad \begin{array}{l} i = 1, 2, \dots, N \\ t = 1965, \dots, 1989 \end{array} \quad (23)$$

which can be viewed as an autoregressive approximation to equation (11c) with  $\rho = 1$ . These equations were estimated both allowing  $g_i$  to differ across countries (Table V) and imposing a common  $g_i = g$  (Table VI). The results of Table V indicate that there are considerable variations in the estimates of  $g_i$ ,  $\gamma_{1i}$ , and  $\gamma_{2i}$  across all country groupings. The values of  $g_i$  are generally precisely estimated, and are similar in magnitude to the estimates of  $g_i$  obtained for the trend-stationary versions of the model presented in Table I, while  $\gamma_{1i}$  and  $\gamma_{2i}$  are usually statistically significant.<sup>14</sup> Comparison of the results in Table V with those in Table VI shows that the imposition of the common growth rate has a considerable effect on the estimates of  $\gamma_{1i}$  and  $\gamma_{2i}$  in the Non-Oil and Intermediate country groupings, but not in the OECD case. A common growth rate is rejected on the demeaned data, LR statistics (95% critical values) are 278.1 (124.3), 143.6 (79.08), and 41.6 (32.67) for the three samples. On the original data, not demeaned, growth rate homogeneity is also rejected in the full and intermediate sample, but not in the OECD, where the LR statistic is 30.25.

Therefore, irrespective of whether there exists a unit root in the output process, the hypothesis of a common technological growth rate across countries is strongly rejected.

#### 4.4. Sigma Convergence

If there is no common steady-state growth rate, then the whole notion of beta convergence including conditional beta convergence has little economic meaning. However, even with heterogeneous growth rates, the model has implications for sigma convergence. Appendix D derives an

<sup>14</sup> The model was also estimated subject to the restrictions that  $H_{0a}: \gamma_{1i} = \gamma_{2i} = 0$ , and  $H_{0b}: \gamma_{2i} = 0$  for all  $i$ . The log likelihood ratio statistics for testing these restrictions in the Non-Oil, Intermediate and OECD countries are, respectively:

$H_{0a}$ : 331.47 (234.0), 222.54 (146.6), and 62.53 (61.66)

$H_{0b}$ : 148.60 (124.3), 95.26 (79.08), and 22.45 (33.92)

where 95% critical values are in parentheses. These tests suggest that at least an AR(2) process is required to adequately capture the dynamics in the output growth series.

Table V. Summary statistics of estimated coefficients from the output growth model in the unit root case (calculated using demeaned data)

	$\hat{g}_i$	$\hat{\gamma}_{1i}$	$\hat{\gamma}_{2i}$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
(LL = 4461.36)				
Mean	0.01768	0.1037	-0.1051	0.05182
(Standard error)	(0.0019)	(0.0285)	(0.0219)	(0.0030)
Median	0.01892	0.0927	-0.1006	0.04344
Standard deviation	0.01905	0.2865	0.2196	0.02992
Minimum	-0.02457	-0.5555	-0.6016	0.01309
Maximum	0.07268	0.9817	0.4063	0.18084
<i>Intermediate group of countries (N = 61)</i>				
(LL = 3089.54)				
Mean	0.02230	0.1640	-0.1113	0.03821
(Standard error)	(0.0021)	(0.0362)	(0.0290)	(0.0026)
Median	0.02156	0.2081	-0.1168	0.03460
Standard deviation	0.01659	0.2827	0.2265	0.01991
Minimum	-0.01789	-0.4206	-0.6235	0.01239
Maximum	0.07256	0.9162	0.4135	0.10282
<i>OECD countries (N = 22)</i>				
(LL = 1411.46)				
Mean	0.02760	0.1534	-0.0117	0.02092
(Standard error)	(0.0018)	(0.0526)	(0.0369)	(0.0016)
Median	0.02692	0.2457	-0.0127	0.02011
Standard deviation	0.00814	0.2412	0.1691	0.00724
Minimum	0.01286	-0.4152	-0.4209	0.01008
Maximum	0.04610	0.5442	0.2749	0.03638

*Notes:*

Estimated coefficients relate to regression equations of the form

$$\Delta(x_{it} - \bar{x}_t) = g_i^d + u_{it}^d \quad \text{where } u_{it}^d = \gamma_{1i}u_{i,t-1}^d + \gamma_{2i}u_{i,t-2}^d + \varepsilon_{it}^d \quad \begin{matrix} i = 1, 2, \dots, N \\ t = 1965, \dots, 1989 \end{matrix}$$

estimated for each country by the exact ML method. Estimates of  $g_i$  are obtained using the formulae  $\hat{g}_i = \hat{g} + \hat{g}_i^d$ , where  $\hat{g}$  is estimated using the aggregate equation  $\Delta\bar{x}_t = \hat{g} + \bar{u}_t$ , where  $\bar{u}_t = \bar{\gamma}_1\bar{u}_{t-1} + \bar{\gamma}_2\bar{u}_{t-2} + \bar{\varepsilon}_t$ , and

$$\bar{x}_t = N^{-1} \sum_{i=1}^N x_{it}$$

See notes to Table I for further details.

equation that describes the evolution of the cross-section variance of output,  $\sigma_t^2$ . The change in variance from an initial value  $\sigma_0^2$  is given by:

$$\sigma_t^2 = \lambda^{2t}\sigma_0^2 + (1 - \lambda^{2t})\sigma_{*0}^2 + \left(\frac{1 - \lambda^{2t}}{1 - \lambda^2}\right)\tau^2 + t^2\sigma_g^2 + 2t\sigma_{0g} \quad (24)$$

where  $\sigma_{*0}^2$  is the variance of the deterministic component of initial output,  $\tau^2$  is the average variance of the time-series (demeaned) disturbances,  $\sigma_g^2$  is the cross-section variance of the growth rate of technology, and  $\sigma_{0g}$  is the covariance between the technology growth rate and initial output. Notice that very similar terms appear in equation (24) to those determining the bias of the Barro regression discussed in Section 3.1.

Table VI. Summary statistics of estimated coefficients from the output growth model in the unit root case under  $H_0: g_i = g$  (calculated using demeaned data)

	$\hat{g}_i$	$\hat{\gamma}_{1i}$	$\hat{\gamma}_{2i}$	$\hat{\tau}_i$
<i>Non-oil countries (N = 102)</i>				
<i>(LL = 4322.31)</i>				
Mean	0.01827	0.2023	-0.0091	0.05424
(Standard error)		(0.0280)	(0.0225)	(0.0030)
Median		0.2432	-0.0036	0.04705
Standard deviation		0.2812	0.2262	0.03033
Minimum		-0.5409	-0.5227	0.01493
Maximum		1.0257	0.5082	0.18222
<i>Intermediate group of countries (N = 61)</i>				
<i>(LL = 3014.75)</i>				
Mean	0.02308	0.2487	-0.0306	0.04012
(Standard error)		(0.0351)	(0.0287)	(0.0027)
Median		0.2664	-0.0411	0.03501
Standard deviation		0.2740	0.2243	0.02071
Minimum		-0.4237	-0.4933	0.01289
Maximum		0.9874	0.4875	0.10599
<i>OECD countries (N = 22)</i>				
<i>(LL = 1390.67)</i>				
Mean	0.02726	0.2169	-0.0606	0.02175
(Standard error)		(0.0541)	(0.0366)	(0.0017)
Median		0.3123	-0.0523	0.02068
Standard deviation		0.2477	0.1678	0.00759
Minimum		-0.2947	-0.3965	0.01010
Maximum		0.5508	0.2872	0.03888

*Notes:*

Estimated coefficients relate to regression equations of the form given in Table V, but estimated by the exact ML method subject to the restriction  $g_i = g$ , for all  $i$ .

See notes to Table I for further details.

The result in equation (24) has a number of features worth noting. First, all the terms are necessarily non-negative except  $\sigma_{0g}$ . Therefore a negative value of this covariance is a necessary condition for sigma convergence. If  $\sigma_g^2 = 0$ ,  $\sigma_i^2$  converges to a constant, which is independent of the initial variance and is equal to  $\sigma_{*0}^2 + \tau^2/(1 - \lambda^2)$ . If  $\sigma_g^2$  is non-zero,  $\sigma_i^2$  will eventually increase, but in the short run could go either up or down. In the long run, the distribution of  $g_i$  determines the time evolution of the cross-section distribution of output. If the distribution of  $g_i$  is bimodal, for example high in rich countries and low in poor countries, this could give rise to the type of bimodal distribution discussed by Quah (1996) and Bianchi (1996).

We used formula (24) to predict the variance of output in the three samples. Figures 1 and 2 plot predicted and sample variances using the estimates in Tables III and IV, respectively.<sup>15</sup> Note that unlike Evans (1996b) who also analyses the variance, these predictions were not derived from the variance data itself, but by the direct application of the formula in equation (24). The predicted values in Figure 1 are based on the estimates in Table III, which do not impose homogeneity in the growth rates, while the predicted values in Figure 2 are computed using the

<sup>15</sup> The relevant estimates of  $\sigma_0^2$ ,  $\sigma_{*0}^2$ ,  $\sigma_g^2$ ,  $\sigma_{0g}$  and  $\tau^2$  used to compute the predicted values of  $\sigma_i^2$  are summarized in Table VII.

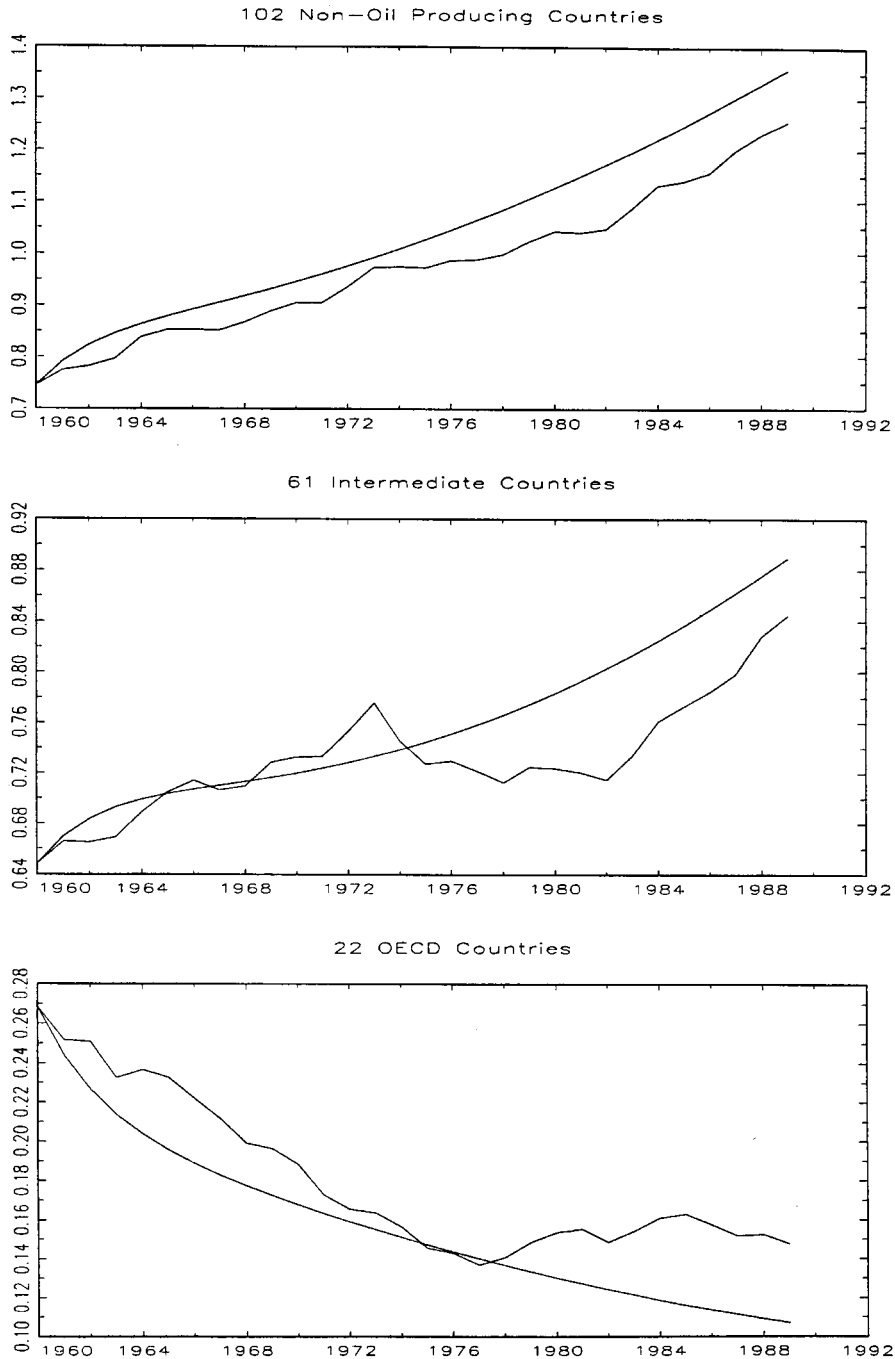


Figure 1. Evolution of cross-section variances of log per capita output (actual sample values and predictions based on Table III). Predicted values are computed using equation (24) and the parameter estimates obtained from the results of Table III

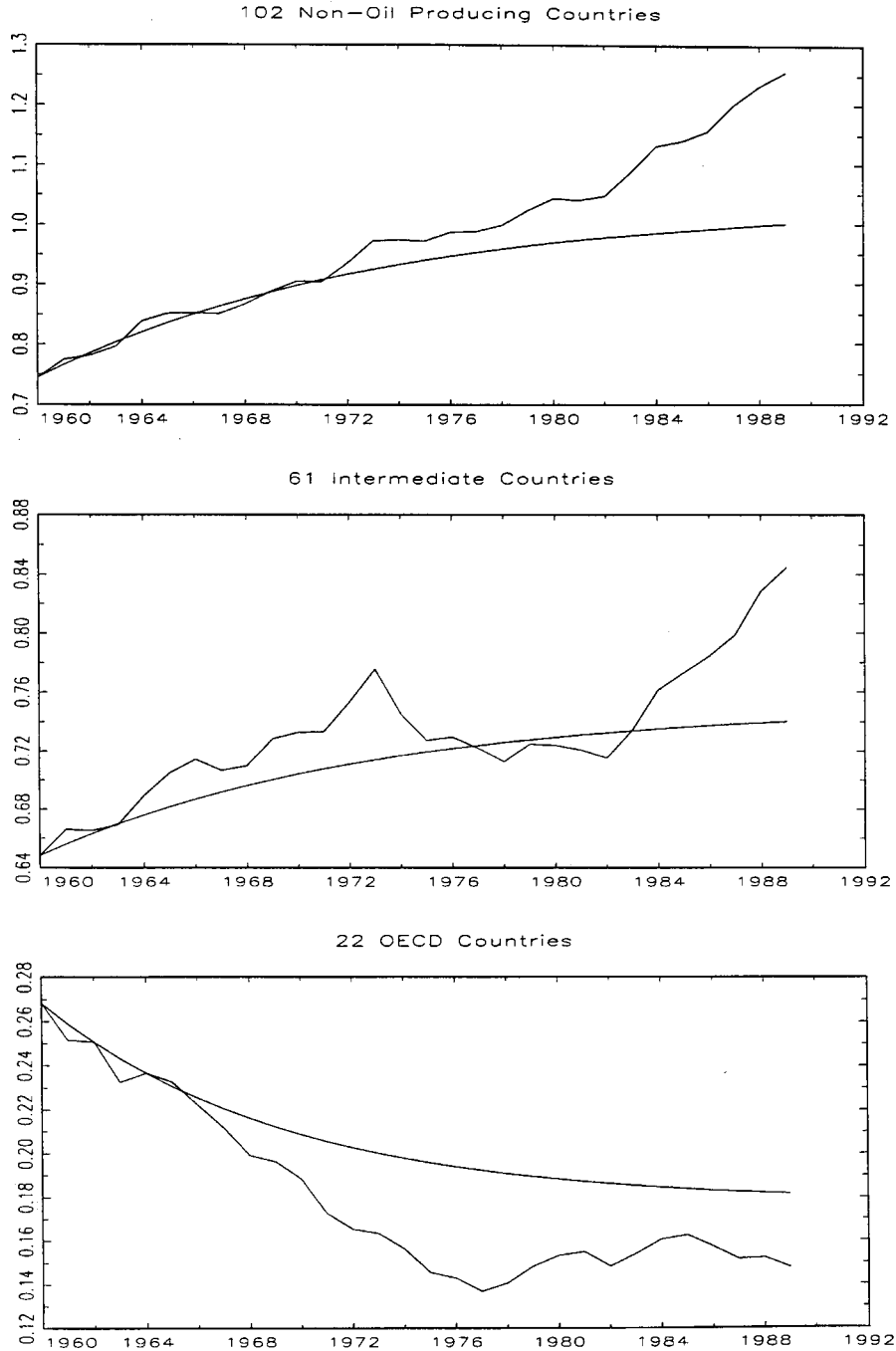


Figure 2. Evolution of cross-section variances of log per capita output (actual sample values and predictions based on Table IV). Predicted values are computed using equation (24) and the parameter estimates obtained from the results of Table IV

Table VII. Estimated cross-country variances and covariances

	Non-oil (102)	Intermediate (61)	OECD (22)
$\hat{\sigma}_0^2$	0.74510	0.64840	0.26820
<i>Estimated statistics based on the results of Table III</i>			
$\hat{\lambda}$	0.77460	0.80960	0.79990
$\hat{\sigma}_{*0}^2$	0.84000	0.71380	0.21470
$\hat{\sigma}^2$	0.00040	0.00029	0.00006
$\hat{\tau}^2$	0.00230	0.00130	0.00039
$\hat{\sigma}_{*g}$	0.00299	-0.00146	-0.00223
$\hat{\sigma}_{0g}$	0.00301	-0.00130	-0.00274
<i>Estimated statistics based on the results of Table IV</i>			
$\hat{\lambda}$	0.96280	0.96140	0.94770
$\hat{\sigma}_{*0}^2$	0.99180	0.72830	0.17290
$\hat{\tau}^2$	0.00297	0.00169	0.00048

Notes:

$$\hat{\sigma}_0^2 = (N - 1)^{-1} \sum_{i=1}^N (x_{i0} - \bar{x}_0)^2$$

is the sample estimate of the cross-section variance of log per capita output in the base year (1960),  $\hat{\sigma}_{*0}^2$  is the sample estimate of the variance of  $x_{i0}^*$ , the deterministic component of the output in the base year,  $\hat{\sigma}_{*g}$  and  $\hat{\sigma}_{0g}$  are the sample estimates of the covariances of  $g_i$  (growth rate of technology) with  $x_{i0}$  and  $x_{i0}^*$ , respectively;

$$\hat{\sigma}_g^2 = (N - 1)^{-1} \sum_{i=1}^N (\hat{g}_i - \hat{g}_0)^2$$

is the sample estimate of the cross-section variance of technology growth rates,  $\hat{\lambda}$  and  $\hat{\tau}^2$  are the ML estimates of  $\lambda$  and  $\tau^2$  from Tables III and IV.

Note that since the estimates in Table IV assume  $g_i = g$ , then  $\sigma_{*g} = \sigma_{0g} = \sigma_g^2 = 0$ .

estimates in Table IV, which do impose homogeneity of growth rates. The plots in Figure 1 match the data rather well, and much better than those in Figure 2. This suggests that growth rate heterogeneity is a major force driving the variance of output. Because growth rate heterogeneity is rather low in the OECD, the predicted OECD variance shows a downward trend over the whole period in Figure 1. However, equation (24) predicts that even in the case of the OECD sample the growth rate heterogeneity (if it persists) will eventually dominate and start causing the predicted variance to rise.

## 5. CONCLUSIONS

The stochastic Solow model provides a tight theoretical framework within which the growth process can be systematically interpreted and has implications which are rather different from the standard deterministic version. The empirical analysis of the logarithm of per-capita output, output for short, in this paper indicates that data for 102 countries over the last 30 years strongly rejects the hypothesis that technology growth rates are equal across countries. While it is well known that growth rates differ, the econometric implications of ignoring growth rate heterogeneity do not seem to have been appreciated. We examine the econometric properties of estimators of beta convergence, as traditionally defined, in cross-section and pooled datasets



and derive their asymptotic biases. When the homogeneity assumption is relaxed, estimated speeds of beta convergence increase from the 2% per annum found in cross-section to an average of around 30% per annum. This changes the mean adjustment lag from 49 years to 2.33 years. However, this coefficient is relatively imprecisely estimated and the interpretation of it as a measure of beta convergence is highly questionable in the context of a stochastic Solow growth model.

Despite the much higher estimates we obtain for the speed of convergence, we cannot reject the null hypothesis of a unit root in the output process. However, failure to reject this null is not evidence against the Solow model. In the stochastic Solow model which we analyse the coefficient of the lagged dependent variable in Augmented Dickey–Fuller type regressions will be  $(1 - \lambda)(1 - \rho)$ , where  $\rho$  is the serial correlation coefficient of the technology process and  $(1 - \lambda) = (1 - \alpha)(n + g + \delta - h)$ , where  $\alpha$  is the capital exponent,  $n$  the rate of population growth,  $g$  the rate of growth of technology,  $\delta$  the depreciation rate, and  $h$  depends on the difference between the deterministic steady-state solution and the mean of the stochastic steady-state distribution of the effective capital labour ratio. Tests for a unit root in the output process thus test whether  $(1 - \rho)(1 - \alpha)(n + g + \delta - h) = 0$ . These tests are likely to lack power, particularly because the stochastic Solow model we analyse suggests that the disturbance in the output equation will have a moving-average form with an MA root close to minus unity. In these circumstances the probability of rejecting the null when the root of the AR part of the output equation is close to unity is likely to be low, despite the more powerful panel tests we employ. Even putting aside the power issue, the unit root hypothesis does not relate directly to the speed of convergence or the capital exponent. Each of the three components  $(1 - \rho)$ ,  $(1 - \alpha)$ , and  $(n + g + \delta - h)$  could have plausible values, e.g. 0.1, 0.66 and 0.08, yet the product of them, 0.005, would be indistinguishable from zero. The conventional interpretation would be valid only if  $\rho = 0$  and  $h = 0$ , neither of which are likely.

In the context of a stochastic model, traditional estimates of beta convergence do not have their usual interpretation. When growth rates of technology differ across countries the speed of convergence is informative only about the within country output movements. However, this growth heterogeneity will be the major determinant of the cross-country output dispersion. Our estimates indicate that growth of technology has been higher in OECD countries with a smaller dispersion as compared to the world as a whole. In consequence, global dispersion is increasing: countries are diverging, not converging.

## APPENDIX A: ASYMPTOTIC BIAS IN THE ‘BARRO CROSS-SECTION REGRESSION’

The Barro cross-section regression is derived by iterating on equation (13) to obtain

$$x_{iT} - x_{i0} = \left( \frac{1 - \lambda^T}{1 - \lambda} \right) \mu_i + (1 - \lambda) \left( \sum_{j=1}^T j \lambda^{T-j} \right) g_i - (1 - \lambda^T) x_{i0} + \xi_{iT} \quad (\text{A1})$$

for  $i = 1, 2, \dots, N$ , where

$$\xi_{iT} = \sum_{j=0}^{T-1} \lambda^j \varepsilon_{i,T-j} \quad (\text{A2})$$

Also from equation (13a)

$$\mu_i = \lambda_i g_i + (1 - \lambda_i) x_{i0}^* \quad (\text{A3})$$

where

$$x_{i0}^* = a_{i0} + \boldsymbol{\psi}' \mathbf{w}_i - \left( \frac{\alpha h_i}{1 - \lambda_i} \right) \quad (\text{A4})$$

$$\boldsymbol{\psi} = \left( \frac{\alpha}{1 - \alpha}, \frac{-\alpha}{1 - \alpha} \right)' \text{ and } \mathbf{w}_i = (\log s_i, \log(n_i + g_i + \delta - h_i))'$$

The Barro regression assumes  $\lambda_i = \lambda$ ,  $i = 1, 2, \dots, N$ , and involves regressing  $x_{iT} - x_{i0}$  on  $x_{i0}$ , and then recovering an estimate of  $\lambda$  from that of  $b_T = \lambda^T - 1$ , the coefficient of  $x_{i0}$ , for a finite  $T$  and a large enough value of  $N$ . In this appendix we consider the properties of the OLS estimator of  $b_T$  for sufficiently large  $N$  and a finite  $T$  under the following assumptions:

*Assumption A1:* The disturbances,  $\varepsilon_{it}$ , follow the two-factor model

$$\varepsilon_{it} = \eta_t + v_{it} \quad i = 1, 2, \dots, N, t = 1, 2, \dots, T \quad (\text{A5})$$

and the country-specific components,  $v_{it}$ , are independently distributed across  $i$ , and follow general stationary processes with zero means and autocovariances  $\gamma_i(s) = 0, 1, 2, \dots$

*Assumption A2:* The various components of  $\mu_i$  (i.e.  $g_i, a_{i0}, w_i, h_i$ ) and  $v_{it}$  are independently distributed both across  $i$  and  $t$ , and are assumed to be generated from a common distribution with mean  $\mu$  and the variance  $\sigma_\mu^2$ . Specifically we assume

$$g_i \sim \text{iid}(g, \sigma_g^2), \quad x_{i0}^* \sim \text{iid}(x_0^*, \sigma_{*0}^2) \quad (\text{A6})$$

$$\mu = \lambda g + (1 - \lambda) x_0^* \quad (\text{A7})$$

and

$$\sigma_\mu^2 = \lambda^2 \sigma_g^2 + (1 - \lambda)^2 \sigma_{*0}^2 + 2\lambda(1 - \lambda) \sigma_{*g} \quad (\text{A8})$$

where  $\sigma_{*g}$  stands for the covariance of  $x_{i0}^*$  and  $g_i$  which in general will be non-zero, unless, of course,  $g_i = g$ .<sup>16</sup>

The estimate of  $b_T$  based on Barro's regression is given by

$$\hat{b}_T = \frac{\sum_{i=1}^N (x_{iT} - x_{i0})(x_{i0} - \bar{x}_0)}{\sum_{i=1}^N (x_{i0} - \bar{x}_0)^2} \quad (\text{A9})$$

<sup>16</sup> Using equation (A4) note that

$$\sigma_{*g} = \text{Cov}(x_{i0}^*, g_i) = \text{Cov}(a_{i0}, g_i) + \boldsymbol{\psi}' \text{Cov}(\mathbf{w}_i, g_i) - \left( \frac{\alpha}{1 - \lambda} \right) \text{Cov}(h_i, g_i)$$

and since  $\mathbf{w}_i$  depends on  $g_i$ , it is unlikely that  $\sigma_{*g}$  will be zero when  $\sigma_g^2 > 0$ .

where

$$\bar{x}_0 = N^{-1} \sum_{i=1}^N x_{i0}$$

Now using equation (A3) in (A1) and noting that

$$(1 - \lambda) \sum_{j=1}^T j \lambda^{T-j} = T - \frac{\lambda(1 - \lambda^T)}{1 - \lambda}$$

we have

$$x_{iT} - x_{i0} = (1 - \lambda^T)x_{i0}^* + Tg - (1 - \lambda^T)x_{i0} + V_{iT} \quad (\text{A10})$$

or

$$= a_T + b_T x_{i0} + V_{iT}$$

where

$$V_{iT} = (1 - \lambda^T)(x_{i0}^* - x_0^*) + T(g_i - g) + \zeta_{iT} \quad (\text{A11})$$

Also using equation (13) to solve for  $x_{i0}$  assuming that the process has started a long time in the past, we have<sup>17</sup>

$$x_{i0} = x_{i0}^* + \sum_{j=0}^{\infty} \lambda^j \varepsilon_{i,-j} \quad (\text{A12})$$

or using equation (A5):

$$x_{i0} - \bar{x}_0 = x_{i0}^* - \bar{x}_0^* + d_{i0} \quad (\text{A13})$$

where

$$d_{i0} = \sum_{j=0}^{\infty} \lambda^j (v_{i,-j} - \bar{v}_{-j})$$

$$\bar{x}_0^* = N^{-1} \sum_{i=1}^N x_{i0}^*, \quad \text{and} \quad \bar{v}_{-j} = N^{-1} \sum_{i=1}^N v_{i,-j} \quad (\text{A14})$$

<sup>17</sup> In deriving this result we have made use of the identity

$$\sum_{j=0}^{\infty} j \lambda^j = \lambda / (1 - \lambda)^2$$

and the fact that  $x_{i0}^* = (\mu_i - \lambda g_i) / (1 - \lambda)$ . See equation (A3) above.

Substituting equation (A10) in (A9) we have

$$\widehat{b}_T - b_T = \frac{\sum_{i=1}^N V_{iT}(x_{i0} - \bar{x}_0)}{\sum_{i=1}^N (x_{i0} - \bar{x}_0)^2}$$

Now using equation (A13) and noting that under Assumption A2,  $v_{it}$  are independently distributed of  $x_{i0}^*$  and  $g_i$ , we obtain

$$\text{Plim}_{N \rightarrow \infty} (\widehat{b}_T) - b_T = \frac{A + B + C}{D} \quad (\text{A15})$$

where

$$\begin{aligned} A &= (1 - \lambda^T) \text{Lim}_{N \rightarrow \infty} \left\{ N^{-1} \sum_{i=1}^N E[(x_{i0}^* - x_0^*)(x_{i0} - \bar{x}_0)] \right\} \\ &= (1 - \lambda^T) \sigma_{*0}^2 \\ B &= T \text{Lim}_{N \rightarrow \infty} \left\{ N^{-1} \sum_{i=1}^N E[(g_i - g)(x_{i0} - \bar{x}_0)] \right\} \\ &= T \sigma_{*g} \\ C &= \text{Lim}_{N \rightarrow \infty} \left\{ N^{-1} \sum_{i=1}^N E(\xi_{iT} d_{i0}) \right\} \end{aligned}$$

and

$$D = \sigma_{*0}^2 + \text{Lim}_{N \rightarrow \infty} \left\{ N^{-1} \sum_{i=1}^N E(d_{i0}^2) \right\}$$

However, using equations (A2) and (A14) it can be shown that

$$E(\xi_{iT} d_{i0}) = \left(1 - \frac{1}{N}\right) \sum_{j=1}^T \sum_{k=0}^{\infty} \lambda^{T-j+k} \gamma_i(j+k)$$

and

$$E(d_{i0}^2) = \left(1 - \frac{1}{N}\right) \left\{ \frac{\gamma_i(0)}{1 - \lambda^2} + 2 \sum_{k=1}^{\infty} \frac{\lambda^k \gamma_i(k)}{1 - \lambda^2} \right\}$$

Substituting the above results in equation (A15) we finally have:

$$\text{Plim}_{N \rightarrow \infty} (\widehat{b}_T) - b_T = \frac{(1 - \lambda^T) \sigma_{*0}^2 + T \sigma_{*g} + \sum_{j=1}^T \sum_{k=0}^{\infty} \lambda^{T-j+k} \bar{\gamma}(j+k)}{\sigma_{*0}^2 + \left(\frac{1}{1 - \lambda^2}\right) \left\{ \bar{\gamma}(0) + 2 \sum_{k=1}^{\infty} \lambda^k \bar{\gamma}(k) \right\}} \quad (\text{A16})$$

where

$$\bar{\gamma}(s) = \lim_{N \rightarrow \infty} \left[ N^{-1} \sum_{i=1}^N \gamma_i(s) \right]$$

It is clear from this result that the estimator of  $b_T$  (and hence that of  $\lambda$ ) obtained from the Barro cross-section regression in general will be inconsistent. The conditions under which  $\widehat{b}_T$  is a consistent estimator of  $b_T$  are very stringent and require (1) homogeneity of growth rates, (2) serially uncorrelated  $v_{it}$ 's, and (3)  $\sigma_{*0}^2 = V(x_{i0}^*) = 0$ . The last restriction is extremely unlikely to be satisfied in practice even if observable components of  $x_{i0}^*$  (namely  $s_i$  and  $n_i$ ) are included in the regression.<sup>18</sup> This is because  $x_{i0}^*$  will continue to depend on the unobservable country-specific initial endowment effects,  $a_{i0}$ , which are unlikely to be identical across  $i$ .

The asymptotic bias expression in equation (A16) simplifies considerably if we follow the literature and assume that the technology growth rates are identical across countries (i.e.  $g_i = g$ ), and  $v_{it}$ 's are serially uncorrelated (i.e.  $\bar{\gamma}(k) = 0$ ,  $k = 1, 2, \dots$ ). Under these restrictions the asymptotic bias is positive for all parameter values and is given by

$$\text{Plim}_{N \rightarrow \infty} (\widehat{b}_T) - b_T = \frac{(1 - \lambda^T) \sigma_{*0}^2}{\sigma_{*0}^2 + \frac{\tau^2}{1 - \lambda^2}} \quad (\text{A17})$$

where  $\tau^2 = \bar{\gamma}(0)$ . In order to get some idea of the quantitative importance of the bias in this case we rewrite equation (A17) as

$$\text{Plim}_{N \rightarrow \infty} (\widehat{\lambda}^T) = \lambda^T + \frac{1 - \lambda^T}{1 + q} = 1 - \frac{q\lambda^T}{1 + q}$$

where

$$q = \frac{\tau^2 / (1 - \lambda^2)}{\sigma_{*0}^2}$$

which is the ratio of the average long-run time-series variance of  $v_{it}$  to the cross-section variance of  $x_{i0}^*$ .<sup>19</sup> In situations where the time-series variance is small relative to the cross-section variance (which seems to be the case in practice),  $q$  will be small and  $\widehat{\lambda}$  will be biased towards unity even if the true value of  $\lambda$  is equal to zero.

#### APPENDIX B: ASYMPTOTIC BIAS OF THE POOLED 'FIXED-EFFECT' ESTIMATOR OF THE CONVERGENCE PARAMETER UNDER HETEROGENEOUS GROWTH RATES

Consider the model

$$x_{it} = \mu_i + \theta_i t + \lambda x_{i,t-1} + \varepsilon_{it} \quad \begin{array}{l} i = 1, 2, \dots, N \\ t = 1, 2, \dots, T \end{array} \quad (\text{B1})$$

<sup>18</sup> The Barro regression with  $s_i$  and  $n_i$  included as additional regressors is known as the 'conditional Barro regression'.

<sup>19</sup> The exact expression for  $\sigma_{*0}^2$  in terms of the structural parameters of the stochastic Solow model can be obtained using equation (A4) and depend on the variance and covariances of  $a_{i0}$ ,  $s_i$ ,  $n_i$ , and  $h_i$  (assuming  $\lambda_i = \lambda$  and  $g_i = g$ ).

where  $\theta_i = g_i(1 - \lambda)$ , and suppose that  $\lambda$  is estimated ignoring the heterogeneity of the growth rates,  $g_i$ , by pooling the time-series processes  $x_{it}$  across  $i$  and treating  $\mu_i$  as unobservable fixed effects. Denoting the pooled ‘fixed-effect’ estimator of  $\lambda$  by  $\hat{\lambda}_{FE}$  we have (see, for example, Pesaran and Smith, 1995),

$$\begin{pmatrix} \hat{\theta}_{FE} \\ \hat{\lambda}_{FE} \end{pmatrix} = \begin{pmatrix} N(\mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T) & \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \mathbf{x}_{i,-1} \\ \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \mathbf{x}_{i,-1} & \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{x}_{i,-1} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \mathbf{x}_i \\ \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{x}_i \end{pmatrix} \tag{B2}$$

where  $\mathbf{H}_T = \mathbf{I}_T - \boldsymbol{\tau}_T(\boldsymbol{\tau}'_T \boldsymbol{\tau}_T)^{-1} \boldsymbol{\tau}_T$ ,  $\boldsymbol{\tau}_T = (1, 1, \dots, 1)'$ ,  $\mathbf{t}_T = (1, 2, \dots, T)'$ ,  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$ , and  $\mathbf{x}_{i,-1} = (x_{i0}, x_{i1}, \dots, x_{i,T-1})'$ . Substituting  $\mathbf{x}_{it}$  from equation (B1) in (B2) yields

$$\begin{pmatrix} \hat{\theta}_{FE} - \bar{\theta} \\ \hat{\lambda}_{FE} - \lambda \end{pmatrix} = \begin{pmatrix} N(\mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T) & \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \mathbf{x}_{i,-1} \\ \sum_{i=1}^N \mathbf{x}'_{i,t-1} \mathbf{H}_T \mathbf{t}_T & \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{x}_{i,-1} \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \boldsymbol{\varepsilon}_i \\ \sum_{i=1}^N (\theta_i - \bar{\theta})(\mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{t}_T) + \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \boldsymbol{\varepsilon}_i \end{pmatrix} \tag{B3}$$

where

$$\bar{\theta} = N^{-1} \sum_{i=1}^N \theta_i$$

and  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT})'$ .

Now using equation (B1) and noting that  $|\lambda| < 1$ , we can solve for  $x_{it}$  as

$$x_{it} = \frac{\mu_i}{1 - \lambda} + \frac{\theta_i}{1 - \lambda} \left( t - \frac{\lambda}{1 - \lambda} \right) + \sum_{j=0}^{\infty} \lambda^j \varepsilon_{i,t-j}$$

or in vector notation (recall that  $g_i = \theta_i/(1 - \lambda)$ ),

$$\mathbf{x}_i = \left( \frac{1}{1 - \lambda} \right) \left( \mu_i - \frac{\lambda \theta_i}{1 - \lambda} \right) \boldsymbol{\tau}_T + g_i \mathbf{t}_T + \sum_{j=0}^{\infty} \lambda^j \boldsymbol{\varepsilon}_{i,-j}$$

where  $\boldsymbol{\varepsilon}_{i,-j} = (\varepsilon_{i,1-j}, \varepsilon_{i,2-j}, \dots, \varepsilon_{i,T-j})'$ . Similarly,

$$\mathbf{x}_{i,-1} = \left( \frac{1}{1 - \lambda} \right) \left( \mu_i - \frac{\theta_i}{1 - \lambda} \right) \boldsymbol{\tau}_T + g_i \mathbf{t}_T + \sum_{j=0}^{\infty} \lambda^j \boldsymbol{\varepsilon}_{i,-j-1}$$

Using this result, we obtain the following expressions for the various sums in equation (B3):

$$\begin{aligned}
N^{-1} \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \boldsymbol{\varepsilon}_i &= \mathbf{t}'_T \mathbf{H}_T \bar{\boldsymbol{\varepsilon}} \\
N^{-1} \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \boldsymbol{\varepsilon}_i &= \mathbf{t}'_T \mathbf{H}_T \left( \frac{\sum_{i=1}^N g_i \boldsymbol{\varepsilon}_i}{N} \right) + \sum_{j=0}^{\infty} \lambda^j \left( \frac{\sum_{i=1}^N \boldsymbol{\varepsilon}'_i \mathbf{H}_T \boldsymbol{\varepsilon}_{i,-j-1}}{N} \right) \\
N^{-1} \sum_{i=1}^N \mathbf{t}'_T \mathbf{H}_T \mathbf{x}_{i,-1} &= \bar{g} (\mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T) + \mathbf{t}'_T \mathbf{H}_T \left( \sum_{j=0}^{\infty} \lambda^j \bar{\boldsymbol{\varepsilon}}_{-j-1} \right) \\
N^{-1} \sum_{i=1}^N \mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{x}_{i,-1} &= \left( \frac{\sum_{i=1}^N g_i^2}{N} \right) (\mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T) \\
&\quad + \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \lambda^{r+s} \left( \frac{\sum_{i=1}^N \boldsymbol{\varepsilon}'_{i,-r-1} \mathbf{H}_T \boldsymbol{\varepsilon}_{i,-s-1}}{N} \right) \\
&\quad + 2 \mathbf{t}'_T \mathbf{H}_T \sum_{j=0}^{\infty} \lambda^j \left( \frac{\sum_{i=1}^N g_i \boldsymbol{\varepsilon}_{i,-j-1}}{N} \right) \\
N^{-1} \sum_{i=1}^N (\theta_i - \bar{\theta}) (\mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{t}_T) &= (1 - \lambda) \left( \frac{\sum_{i=1}^N (g_i - \bar{g})^2}{N} \right) \mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T \\
&\quad + \mathbf{t}'_T \mathbf{H}_T \left( \sum_{j=0}^{\infty} \lambda^j \bar{\boldsymbol{\varepsilon}}_{-j-1} \right) \\
N^{-1} \sum_{i=1}^N (\theta_i - \bar{\theta}) (\mathbf{x}'_{i,-1} \mathbf{H}_T \mathbf{t}_T) &= (1 - \lambda) \left( \frac{\sum_{i=1}^N (g_i - \bar{g})^2}{N} \right) \mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T \\
&\quad + \mathbf{t}'_T \mathbf{H}_T \left( \sum_{j=0}^{\infty} \lambda^j \bar{\boldsymbol{\varepsilon}}_{-j-1} \right)
\end{aligned}$$

where

$$\bar{g} = N^{-1} \sum_{i=1}^N g_i, \quad \text{and} \quad \bar{\boldsymbol{\varepsilon}}_{-j} = N^{-1} \sum_{i=1}^N \boldsymbol{\varepsilon}_{i,-j}$$

(with  $\bar{\boldsymbol{\varepsilon}}_0 = \bar{\boldsymbol{\varepsilon}}$ ). The dominant term of these summations is  $\mathbf{t}'_T \mathbf{H}_T \mathbf{t}_T$  which is of order  $T^3$ . Hence, for large enough  $T$ , and assuming that

$$N^{-1} \sum_{i=1}^N (g_i - \bar{g})^2 > 0,$$

we have

$$\text{Plim}_{T \rightarrow \infty} \begin{pmatrix} \hat{\theta}_{\text{FE}} - \bar{\theta} \\ \hat{\lambda}_{\text{FE}} - \lambda \end{pmatrix} = \begin{bmatrix} 1 & \bar{g} \\ \sum_{i=1}^N g_i^2 & \bar{g} \\ \bar{g} & \frac{\sum_{i=1}^N g_i^2}{N} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \sum_{i=1}^N (g_i - \bar{g})^2 \\ (1 - \lambda) \frac{\sum_{i=1}^N (g_i - \bar{g})^2}{N} \end{bmatrix}$$

and hence, noting that  $\bar{\theta} = \bar{g}(1 - \lambda)$ ,

$$\text{Plim}_{T \rightarrow \infty} (\hat{\theta}_{\text{FE}}) = 0, \quad \text{Plim}_{T \rightarrow \infty} (\hat{\lambda}_{\text{FE}}) = 1 \tag{B4}$$

Note that these results hold for any finite  $N$ , so long as

$$N^{-1} \sum_{i=1}^N (g_i - \bar{g})^2 > 0$$

In the case where  $\varepsilon_{it}$  has a common component as specified in equation (18), then this can be removed by demeaning and equation (B1) takes the form of equation (20). Using derivations along the lines given above, it can be shown that variation in the  $g_i$  will cause bias. But to obtain equation (B4) we need  $N$  sufficiently large so that  $\text{Cov}(\varepsilon_{it} - \bar{\varepsilon}_t, \varepsilon_{jt} - \bar{\varepsilon}_t) \approx 0$ , for  $i \neq j$ .

APPENDIX C: EXACT MAXIMUM LIKELIHOOD ESTIMATION OF SOLOW GROWTH MODEL UNDER THE NULL HYPOTHESIS,  $H_0: g_i = g, \forall i$

The output process under the Solow model is given by equation (13) or, equivalently, by equation (14):

$$\begin{aligned} x_{it} &= c_i + g_i t + u_{it} \\ u_{it} &= \lambda_i u_{i,t-1} + \varepsilon_{it} \end{aligned}$$

Here for the purpose of maximum likelihood (ML) estimation we also assume that  $\varepsilon_{it} \sim N(0, \tau_i^2)$ . The procedure described below can also be applied to the demeaned data.

Under the assumption of Solow's model where  $|\lambda_i| < 1$ , and assuming the output processes began a long time before the start of the observations in 1960, the exact log-likelihood function for the  $i$ th country output process is given by

$$l_i(\boldsymbol{\theta}_i) = -\frac{T}{2} \log(2\pi\tau_i^2) + \frac{1}{2} \log(1 - \lambda_i^2) - \frac{1}{2\tau_i^2} \boldsymbol{\varepsilon}'_i \mathbf{R}(\lambda_i) \boldsymbol{\varepsilon}_i \tag{C1}$$



where

$$\begin{aligned} \varepsilon_i &= \mathbf{x}_i - c_i \boldsymbol{\tau}_T - g_i \mathbf{t}_T \\ \mathbf{x}_i &= (x_{i1}, x_{i2}, \dots, x_{iT})' \\ \boldsymbol{\tau}_T &= (1, 1, \dots, 1)', \quad \mathbf{t}_T = (1, 2, \dots, T)' \\ \mathbf{R}(\lambda_i) &= \begin{bmatrix} 1 & -\lambda_i & 0 & 0 & \dots & 0 \\ -\lambda_i & 1 + \lambda_i^2 & -\lambda_i & 0 & \dots & 0 \\ 0 & & & & & \\ \vdots & & & & & \\ 0 & \dots & \dots & -\lambda_i & 1 + \lambda_i^2 & -\lambda_i \\ 0 & \dots & \dots & & -\lambda_i & 1 \end{bmatrix} \end{aligned}$$

and  $\boldsymbol{\theta}_i = (c_i, g_i, \lambda_i, \tau_i^2)'$ , for  $i = 1, 2, \dots, N$ . Moreover, since  $\varepsilon_{it}$ 's are assumed to be independently distributed across  $i$ , the log-likelihood function for the whole system of output equations is given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^N l_i(\boldsymbol{\theta}_i) \quad (\text{C2})$$

where  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_N)'$ .

In the unrestricted case where it is not required for  $g_i$  to be the same across countries, the exact ML estimators of  $\boldsymbol{\theta}_i$  can be obtained by application of standard algorithms to the log-likelihood function of each country separately. (See, for example, Pesaran and Pesaran, 1997.)

Under the restrictions  $H_0: g_i = g$ , the exact ML estimators of the Solow model for different countries can be obtained by maximizing  $l(\boldsymbol{\theta})$  directly. We first note that, for a given  $g = g^{(j)}$ , the ML estimators of  $c_i$ ,  $\lambda_i$ , and  $\tau_i^2$  are given by the regression of  $[x_{it} - g^{(j)}t]$  on an intercept, where the disturbances are assumed to follow an AR(1) process, estimated using the exact ML method (see above). This provides estimates of  $c_i$ ,  $\lambda_i$ ,  $\tau_i^2$  which we denote by  $\hat{g}(g^{(j)})$ ,  $\hat{\lambda}_i(g^{(j)})$ , and  $\hat{\tau}_i(g^{(j)})$ , respectively. To estimate  $g$  we note that

$$\frac{\partial l(\boldsymbol{\theta})}{\partial g} = \frac{1}{2} \sum_{i=1}^N \left\{ \frac{\mathbf{t}'_T \mathbf{R}(\lambda_i) (\mathbf{x}_i - \delta_i \boldsymbol{\tau}_T - g \mathbf{t}_T)}{\tau_i^2} \right\}$$

and for given values of  $\lambda_i$ ,  $c_i$ , and  $\tau_i^2$ ,  $i = 1, 2, \dots, N$ , the estimator of  $g$  is given by

$$\hat{g}(\boldsymbol{\psi}_i) = \frac{\sum_{i=1}^N \left\{ \frac{\mathbf{t}'_T \mathbf{R}(\lambda_i) \mathbf{x}_i - \delta_i \mathbf{t}'_T \mathbf{R}(\lambda_i) \boldsymbol{\tau}_T}{\tau_i^2} \right\}}{\sum_{i=1}^N \left\{ \frac{\mathbf{t}'_T \mathbf{R}(\lambda_i) \mathbf{t}_T}{\tau_i^2} \right\}} \quad (\text{C3})$$

where  $\boldsymbol{\psi}_i = (\lambda_i, c_i, \tau_i^2)'$ . The above results suggest the following iterative numerical procedure for estimating the parameters of the cross-country output processes under  $g_i = g, \forall i$ :

*Step 1:* Fix  $g = \hat{g}^{(1)}$ , where  $\hat{g}^{(1)}$  is the mean value of the estimates of  $g_i$  over  $i = 1, 2, \dots, N$  obtained under the unrestricted model.

*Step 2:* Estimate the regression of  $\{x_{it} - \hat{g}^{(1)}\}$  on an intercept, assuming the disturbances have an AR(1) process, and thereby obtain estimated values of  $\lambda_i$ ,  $c_i$ , and  $\tau_i^2$  denoted  $\hat{\lambda}_i(\hat{g}^{(1)})$ ,  $\hat{c}_i(\hat{g}^{(1)})$ , and  $\hat{\tau}_i(\hat{g}^{(1)})$ , respectively. Calculate the system-wide value of the log-likelihood function,  $l(\hat{\theta}^{(1)})$ .

*Step 3:* Use equation (C3) to obtain an estimate of  $g$ , denoted by  $\hat{g}^{(2)}$ , based on the estimated values  $\hat{\lambda}_i(\hat{g}^{(1)})$ ,  $\hat{c}_i(\hat{g}^{(1)})$  and  $\hat{\tau}_i(\hat{g}^{(1)})$  in Step 2.

*Step 4:* Repeat Steps 1 and 2 with  $\hat{g}^{(j)}$  replaced by  $\hat{g}^{(j+1)}$  for  $j = 1, 2, \dots$ , until the differences between successive estimates of  $\theta$  and the differences in the maximized log-likelihood values in successive iterations are sufficiently small.

#### APPENDIX D: EVOLUTION OF CROSS-COUNTRY DISPERSION OF PER CAPITA OUTPUT IN THE SOLOW MODEL

In this appendix we examine the implications of the stochastic Solow model for sigma convergence. We follow the assumptions and definitions set out in Appendix A, and also assume that  $\lambda_i = \lambda$ . Then, using equation (A1) and noting that

$$(1 - \lambda) \sum_{j=1}^T j\lambda^{T-j} + \lambda(1 - \lambda^T)/(1 - \lambda) = T$$

it is easily seen that

$$x_{iT} = \lambda^T x_{i0} + (1 - \lambda^T)x_{i0}^* + Tg_i + \zeta_{iT} \quad (\text{D1})$$

while a demeaned version of equation (D1) is given by

$$(x_{iT} - \bar{x}_T) = \lambda^T(x_{i0} - \bar{x}_0) + (1 - \lambda^T)(x_{i0}^* - \bar{x}_0^*) + T(g_i - \bar{g}) + d_{iT} \quad (\text{D2})$$

where

$$d_{iT} = \zeta_{iT} - \bar{\zeta}_T = \sum_{j=0}^{T-1} \lambda^j (v_{i,T-j} - \bar{v}_{T-j})$$

Now, squaring, averaging across  $i$ , and taking probability limits of both sides of equation (D2) as  $N \rightarrow \infty$ , and noting that under Assumption A1,  $v_{it}$  is distributed independently of  $x_{i0}$  and  $x_{i0}^*$ , we obtain

$$\begin{aligned} \sigma_T^2 &= \lambda^{2T} \sigma_0^2 + (1 - \lambda^T)^2 \sigma_{*0}^2 + T^2 \sigma_g^2 + \sigma_{dT}^2 \\ &\quad + 2\lambda^T(1 - \lambda^T) \text{Cov}(x_{i0}, x_{i0}^*) + 2T\lambda^T \sigma_{0g} + 2T(1 - \lambda^T) \sigma_{*g} \end{aligned} \quad (\text{D3})$$

where<sup>20</sup>

$$\begin{aligned} \sigma_T^2 &= \text{Var}(x_{iT}) = \text{Plim}_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N (x_{iT} - \bar{x}_T)^2 \right] \\ \sigma_{dT}^2 &= \text{Var}(d_{iT}) = \text{Plim}_{N \rightarrow \infty} \left[ \frac{1}{N} \sum_{i=1}^N (\zeta_{iT} - \bar{\zeta}_T)^2 \right] = \tau^2 \left( \frac{1 - \lambda^{2T}}{1 - \lambda^2} \right) \end{aligned} \quad (\text{D4})$$

<sup>20</sup> For expositional simplicity the expression for  $\sigma_{dT}^2$  in equation (D4) is derived assuming that  $v_{it}$ 's are not serially correlated.

$\tau^2 = \text{Var}(v_{it})$ , and

$$\tau^2 = \lim_{N \rightarrow \infty} \left( \frac{1}{N} \sum_{i=1}^N \tau_i^2 \right) \quad (\text{D5})$$

It is worth noting here that  $\sigma_T^2$  is the cross-section variance of (log) per capita output at time  $T$ ,  $\sigma_{dT}^2$  is the variance of the  $d_{iT}$ , and  $\sigma_{*0}^2$  is the cross-country variance of the deterministic components of the initial output,  $x_{i0}^*$ , and  $\sigma_{0g}$  (respectively,  $\sigma_{*g}$ ) is the cross-country covariance between  $x_{i0}$  (respectively,  $x_{i0}^*$ ) and  $g_i$ . Also using equation (A13) we have

$$\sigma_{0g} = \sigma_{*g} \quad \text{and} \quad \text{Cov}(x_{i0}, x_{i0}^*) = \text{Var}(x_{i0}^*) = \sigma_{*0}^2 \quad (\text{D6})$$

Finally, using equations (D4) and (D6) in (D3) and simplifying, we have

$$\sigma_T^2 = \lambda^{2T} \sigma_0^2 + (1 - \lambda^{2T}) \sigma_{*0}^2 + \left( \frac{1 - \lambda^{2T}}{1 - \lambda^2} \right) \tau^2 + T^2 \sigma_g^2 + 2T \sigma_{0g} \quad (\text{D7})$$

which describes the evolution of cross-country variance of output in terms of the initial cross-country variances of initial output ( $x_{i0}$ ), its deterministic component ( $x_{i0}^*$ ), the dispersion in the growth rates of technology ( $g_i$ ), and the covariance of the initial output and  $g_i$ .

#### ACKNOWLEDGEMENTS

Earlier versions of this paper with the title ‘Growth and convergence: a multi-country empirical analysis of the Solow growth model’ were presented at the Second Annual Conference of the ESRC Initiative on the Analysis of Large and Complex Datasets, University of Nottingham, the Annual Conference of the ESRC Development Economics Study Group, University of Leicester, the Sixth International Panel Data Conference, Amsterdam, the Annual Conference of the Royal Economic Society, University of Wales, Swansea, the Annual Conference of the Irish Economic Association, Limerick, and at seminars at Ohio State University, University of California at Los Angeles, San Diego, and Riverside, Birkbeck College, London, and the Universities of Edinburgh, Loughborough and Reading. We are grateful to Steven Durlauf, Michael Binder, and two anonymous referees who provided detailed comments and to Paul Evans for providing copies of his papers; all of whom helped substantially in the revision of this paper. The first two authors gratefully acknowledge financial support from the ESRC (Grant no. H519 25 5003) and from the Isaac Newton Trust of Trinity College, Cambridge.

#### REFERENCES

- Baltagi, B. H. (1995), *Econometric Analysis of Panel Data*, Wiley, Chichester.  
 Barro, R. and X. Sala-i-Martin (1995), *Economic Growth*, McGraw-Hill, New York.  
 Bernard, A. B. and S. Durlauf (1995), ‘Convergence in international output’, *Journal of Applied Econometrics*, **10**, 97–108.  
 Bernard, A. B. and S. Durlauf (1996), ‘Interpreting tests of the convergence hypothesis’, *Journal of Econometrics*, **71**, 161–73.  
 Bianchi, M. (1996), ‘Testing for convergence. Evidence from non-parametric multimodality tests’, mimeo, Bank of England.  
 Binder, M. and M. H. Pesaran (1996), ‘Stochastic growth’, Department of Economics, University of Maryland, Working Paper No. 96-118.

- Canova, F. and A. Marcet (1995), 'The poor stay poor: non-convergence across countries and regions', Paper presented at CEPR Workshop on Empirical Macroeconomics, 1995.
- Durlauf, S. (1996), 'On convergence and divergence of growth rates: an introduction', *Economic Journal*, **106**, July, 1016–18.
- Evans, P. and G. Karras (1996), 'Convergence revisited', *Journal of Monetary Economics*, forthcoming.
- Evans, P. (1996a), 'How fast do economies converge?' *Review of Economics and Statistics*, forthcoming.
- Evans, P. (1996b), 'Using cross-country variances to evaluate growth theories', *Journal of Economic Dynamics and Control*, forthcoming.
- Friedman, M. (1992), 'Do old fallacies ever die?' *Journal of Economic Literature*, **30** (December), 2129–32.
- Heston, A. and R. Summers (1996), 'International price and quantity comparisons: potentials and pitfalls', *American Economic Review*, **86**, 20–24.
- Im, K. S., M. H. Pesaran and Y. Shin (1995), 'Testing for unit roots in dynamic heterogeneous panels', DAE Working Paper, Cambridge. Revised July 1996.
- Islam, N. (1995), 'Growth empirics: a panel data approach', *Quarterly Journal of Economics*, **110**, 4, 1127–1170.
- Kwiatkowski, D., P. C. Phillips, P. Schmidt and Y. Shin (1992), 'Testing the null hypothesis of stationarity against the alternative of a unit root', *Journal of Econometrics*, **54**, 159–78.
- Mankiw, N. G., D. Romer and D. N. Weil (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, **107** (May), 407–37.
- Miller, R. (1995), 'Time-series estimation of convergence rates', mimeo, Department of Economics, Columbia University.
- Nerlove, M. (1996), 'Growth rate convergence, fact or artifact?' mimeo, Department of Agricultural and Resource Economics, University of Maryland.
- Pesaran, M. H. and B. Pesaran (1997), *Working with Microfit 4.0: Interactive Econometric Analysis*, Oxford University Press, Oxford.
- Pesaran, M. H. and L. J. Slater (1980), *Dynamic Regression: Theory and Algorithms*, Ellis Horwood, Chichester.
- Pesaran, M. H. and R. Smith (1995), 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics*, **68**, 79–113.
- Pesaran, M. H., R. Smith and K. S. Im (1996), 'Dynamic linear models for heterogeneous panels', in L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data*, 2nd edition, Kluwer Academic Publishers, Amsterdam.
- Quah, D. (1993), 'Galton's fallacy and tests of the convergence hypothesis', *Scandinavian Journal of Economics*, **95**, 427–43.
- Quah, D. (1996), 'Growth and convergence in models of distribution dynamics', *Economic Journal*, **106**, 1045–55.
- Summers, R. and A. Heston (1991), 'The Penn World Table (Mark 5): An expanded set of international comparisons 1950–1988', *Quarterly Journal of Economics*, **106**, 327–68.