Econometric aspects of convergence: a survey

Ron P. Smith*
Birkbeck, University of London

January 30, 2024

Abstract

The literature on convergence in per-capita income across countries has not converged on a common concept of convergence. It may be within a country towards its own steady state or between countries. Between country convergence may be absolute convergence to the same steady state; conditional convergence to country specific steady states, functions of observed variables; or club convergence to different steady states. It may be measured by beta convergence; sigma convergence; or the presence of a common trend. This paper surveys the econometric issues involved in estimating the rate of convergence; testing for convergence; and specifying the unobserved steady state. The survey suggests that rather than there being different ways to measure a single concept, convergence, the different measures are measuring different things.

JEL Classifications: C1, C33, E10, F43, O4,
Key Words: Econometrics of growth**,** Economic convergence, beta convergence, sigma convergence.

*Corresponding author, email r.smith@bbk.ac.uk

1 Introduction

Neo-classical growth models predict that when countries are similar with respect to preferences and technology, poor countries with low ratios of capital to labour will have high marginal products of capital and thereby should tend to grow at higher rates than rich countries, Barro (1991). Even if technology differs initially, the low income followers can copy the frontier technology from the high income leaders. Rates of return should be higher in poor countries, since capital is scarce. Thus capital should tend to flow to poor countries, as historically labour tended to flow to rich countries. As a result, the per-capita outputs of different countries should converge. The poorer countries should grow faster than the rich at first, as they adopt and invest in best practice technology, then growth slows as they reach the frontier.

While there are examples of this happening in particular countries, there are many complications, particularly the role of growth enhancing or growth inhibiting institutions. Johnson and Papageorgiou, JP, (2020, p165) after an extensive survey conclude "there is a broad consensus of no evidence supporting absolute convergence in cross-country per capita incomes — that is poor countries do not seem to be unconditionally catching up to rich ones." However, interest in the convergence hypothesis was reignited when Patel, Sandefur and Subramanian (2021) and Kremer, Willis and You, KWY, (2021) provided new evidence which suggested that there was a pattern of absolute, or unconditional, convergence across countries since 2000. This reflected both slower growth by richer countries and faster growth by poorer ones. KWY argue that the difference in conclusions resulted from JP considering convergence from a fixed base date 1960, whereas KWY looked at a moving window, allowing the convergence rate to change through time.

As JP p165 note "convergence is hard to pin down, first because the concept can be operationalized in many ways and second, because econometric approaches and data measurement issues remain a challenge in empirical tests of convergence." This paper provides a survey of some of the ways the concept has been operationalised and of the econometric issues involved in estimating the speed of convergence and testing for non-convergence. Although most of the material in this survey is well established in the various literatures, some of it may not be well known, so a fairly basic review may be useful. The survey suggests that rather than there being different ways to measure a single thing, convergence, the different measures are measuring different things. Similarly, rather than there being a single parameter of interest, the speed of adjustment to steady state, quite different parameters are being measured by the different procedures. Provided the relevant moments exist any estimator is a consistent estimator of something, the issue is what that something is. Econometric theory derivations may not be helpful because they are conditional on some assumed data generating process, DGP, which may not be the appropriate one.

Section 2 provides an overview of the terminology; section 3 considers unconditional convergence; section 4 convergence conditional on observable covariates; section 5 intercept heterogeneity and the role of fixed effects in growth panels; section 6 slope heterogeneity; and section 7 contains some concluding comments.

2 Overview

To provide background, the different notions of convergence will be briefly introduced in this section. They will be discussed in more detail in later sections. This survey is about the literature on convergence in the per-capita incomes of countries. There are a range of other literatures which ask very similar questions, using very similar techniques, about the relationship between the growth and size of, for instance, firms, cities or individual incomes. For instance Weill (2013) applies the beta and sigma convergence tests for panel data, discussed below, to measure convergence of bank competition in EU countries. Sutton (1997) surveys Gibrat's Law, which states that the size and growth of firms are independent, so there is no convergence. In these literatures where the number of observations are much larger, the question is often about the form and evolution of the cross-section distribution. While many of the same statistical issues arise, the context can be different. For instance, in the recent past, mergers and acquisitions have tended to be more common among firms than among countries.

Convergence may be within a country as it moves towards its own steady state or between countries as they move towards each other. If all countries are converging to the same unique, globally stable, steady state equilibrium, this is labelled absolute, or unconditional, convergence. With unconditional convergence, initial conditions do not matter, everyone ends up in the same place. Conditional convergence occurs when there is a common speed of convergence to country specific steady states which are a function of observed determinants like investment rates. Club convergence occurs when there may be multiple local steady state equilibria, each with its own basin of attraction, to form a particular club. This can give rise to a bimodal, twin peaks, or a multi-modal distribution of per-capita outputs. For instance, while one group of countries grow, another group may be stuck in a poverty trap from which they cannot escape. Galor (1996) argues that club convergence is consistent not only with the neoclassical paradigm in general but also with constant returns to scale and diminishing marginal productivity in particular. Müller, Stock and Watson (2022) find evidence of convergence clubs.

Thus the steady states for the logarithm of percapita output to which countries are converging may be identical (unconditional convergence with the same steady state growth rate), parallel straight lines (conditional convergence with the same steady state growth rate but different levels), or unrelated (conditional convergence with different growth rates). We will focus on convergence of log per-capita income or output, but one may be interested in convergence of growth rates, or convergence in other variables, such as poverty, life expectancy or human capital. Rodrik (2013) finds unconditional convergence in labour productivity across manufacturing industries in 118 countries, in recent decades. This may be because manufacturing technology is easier to transfer between countries than other sorts of technology.

Beta convergence occurs when, in a regression of growth on previous income, and perhaps other variables, the coefficient (beta) on previous income is negative. This just says that countries that were initially rich grow slower. Sigma convergence occurs when the variance of log per-capita output across countries, declines through time. KWY present graphs showing the variance increasing then falling in recent years. As is shown below, beta convergence is a necessary but not sufficient condition for sigma convergence. If there is club convergence, where countries are converging to two or more steady states, the variance may not decline. The variance may be sensitive to outliers, most countries converging and a few diverging. In these circumstances one may want to use other measures such as the inter-quartile range, or want to describe the dynamics of the whole distribution, the evolution of the shape over time as in Quah (1996) and Bianchi (1997) and many subsequent papers. JP Figure 5 present graphs of the distribution of log per capita GDP in 1960 and 2010.[1] A third notion of convergence treats log per-capita output as an integrated variable and asks whether two countries share a common deterministic and/or stochastic trend. This was introduced by Bernard and Durlauf (1995, 1996) and implies that the distance between the per-capita outputs of two countries declines through time and the long horizon forecast of the expected difference is zero: there is pairwise convergence. Pesaran (2007a) considers pairwise convergence between all $N(N-1)/2$ possible pairs from N countries.

Müller, Stock, and Watson (2022 p858) highlight five features of the data apparent from a panel of 113 countries, 1900-2017, which echo previous findings. There is "a common growth factor, persistent changes in long-term growth rates within countries, a temporally stable dispersion of the historical cross-sectional distribution, extremely persistent country-specific effects, and a possible group structure of cross-country correlations."

A large part of the literature has considered a linear adjustment process by which the logarithm of per-capita GDP in country $i = 1, 2, ...N$ in time period $t = 1, 2, ..., T$, y_{it} converges to its is its steady state value, y^*_{it}

$$\Delta y_{it} = \beta_{it}(y^*_{it} - y_{i,t-1}) + u_{it} \tag{1}$$

where u_{it} is a mean zero disturbance, and β_{it} is a speed of adjustment or rate of convergence, which measures how fast the country catches up.[2] If $t = 1, 2$ it is purely a cross section, otherwise it is a panel. Speeds of adjustment play an important role in many other literatures in economics and finance and the issues discussed in this survey appear in these other literatures. Examples are the slow adjustment to purchasing power parity, Pesaran et al. (2009), and how firms adjust their capital structure and leverage, Westerlund et al. (2022).

Although (1) is the dominant representation, it is not the only one. Phillips and

[1]The focus of the distributional literature has been on whether there is bi-modality, "twin-peaks". However, the 2010 distribution presented by JP appears to show three modes, though this may be a product of the smoothing.

[2]For convenience, the intercept has here been included in the steady state, but this means that if $\beta_{it} = 0$, the expected growth rate is zero, which is not a desirable feature.

Sul (2009) treat the time series representation differently, with a fixed initial income, a time varying technology $A_{it} = A_{i0}\exp(g_{it}t)$ as and where β_{it} is a time varying speed of convergence:

$$y_{it} = y_i^* + (y_{i0} - y_i^*)e^{\beta_{it}t} + g_{it}t.$$

Müller, Stock, and Watson (2022) use low frequency projections combined with a linear factor structure. It is designed for long term forecasting and they discuss the relation between their model and the ones common in the literature. They conclude (p875) "it is clear that there is a wide range of rates of convergence, with some countries having convergence half-lives of less than a century and others having half-lives so long that in a century-long sample, there is essentially no convergence at all."

Within the context of (1), despite the evidence for heterogeneity in β_{it}, much of the literature has supposed that there is a constant speed of convergence β. If, in addition, the frontier was given by US log per capita GDP, $y_{it}^* = a_i + y_{us,t}$, then the estimated equation would just be a regression of growth on an intercept and the gap from the US:

$$\Delta y_{it} = \beta a_i + \beta(y_{us,t} - y_{i,t-1}) + u_{it}$$

with steady state $Y_{it}^* = A_i Y_{us,t}$, where $A_i = \exp(a_i)$.

Within the context of equation (1), there are a number of questions. Firstly, what is the relevant time period. The time unit, t, could represent a year, a 5 year period, a decade or a century? The size of the estimated β_{it} will reflect the length of the time period, though it is usually converted back to a per-annum rate. The argument for using time averaged data over five or ten year periods, rather than annual data, is that it will remove short term cyclical effects. On the other a lot of observations are lost. The benefit of time averaging remains an open question. Secondly, how to specify y_{it}^*, the steady state attractor? This is an unobserved variable. Thirdly, how to estimate β_{it}, the speed of convergence and test the null hypothesis of no convergence? Fourthly, how much homogeneity is there over time? This is an issue of the constancy of parameters over time. JP Figures 7a and 7b, show that there is little predictability in the growth rates, particularly for low income countries, between decades, suggesting parameter instability and raising questions about the existence of a steady state growth rate. If the rate of convergence is constant over time, $\beta_{it} = \beta_i$, and the steady state is a country specific trend output, $y_{it}^* = a_i + g_i t$,

$$\Delta y_{it} = \beta a_i + \beta g_i t - \beta_i y_{i,t-1} + u_{it} \qquad (2)$$

then testing $\beta_i = 0$ is the time series unit root problem for each country. With the addition of lagged changes, $\Delta y_{i,t-i}$, to deal with serial correlation in u_{it} equation (2) takes the form of the augmented Dickey Fuller equation. Under the null of $\beta_i = 0$ in (2), y_{it} is integrated of order one, $I(1)$, a random walk if u_{it} is not serially correlated. The final question is how much homogeneity is there over countries? This is an issue of the constancy of parameters over countries and whether all countries have the same steady state and speed of convergence. The answers to these difficult questions will inform the choice of model and estimator. Below, unless it is specified otherwise, the estimator is assumed to be ordinary least squares applied to the cross section, panel or time series model specified.

Although we will work within this framework, adopting this model is not innocuous, Lee, Pesaran and Smith (1997) show that within the context of a stochastic Solow model the adjustment process is more complex and the expected value of a stochastic steady state will not be equal to a deterministic steady state. Output will have the same time series properties as technology, so if technology has a unit root output will also. In addition there is likely to be a moving average component in u_{it}, which causes a range of complications for estimation and inference on β. The linearisation used to specify the adjustment process will only provide a good approximation close to steady state. Countries may transition through several stages, join one club, before moving on to another club. Growth is episodic, with big breaks for many countries, so smooth convergence to a steady state may not be the appropriate framework.

There are also many measurement issues with the observed data, in particular how to measure constant price GDP per capita in a common currency. But although they are of central importance, they are not the focus of this survey. The parameters of all the statistical models we consider are functions of deeper parameters of structural models, such as the Solow growth model, and an alternative way of estimating speeds of convergence is to calibrate it from these deeper parameters. For instance Fernández-Villaverde, Ohanian, and Yao (2023) use Chinese data to calibrate a neo-classical model with catch-up to US total factor productivity, TFP. They note that Chinese growth matches that of the other East Asian economies that grew quickly. Their estimates imply a catch up rate of 8% and a final total factor productivity, TFP, of 46.7% of the U.S. level. They interpret this as indicating serious structural limitations in China's long-run productivity.

The structure adopted below follows the historical development and moves from the cross section, between country, to the time series, within country, dimension and as Bernard and Durlauf (1996 p163) note: "these results illustrate how the cross-section and time series approaches to convergence make different assumptions both about what one means by convergence and about the properties of the economies under study, and therefore how tests within the two frameworks can lead to very different conclusions concerning cross-country output relationships." Durlauf (2009) discusses the interaction of theory and empirical work in the evolution of the growth literature.

3 Unconditional convergence

While the classical economists were much concerned with growth, Baumol (1986) quotes Marx and Engels, modern growth theory can be dated from Harrod (1939). There followed a lot of work in the 1950s and early 1960s, including Solow (1956) and Swan (1956). Hahn and Mathews (1964) provide a survey. Then from the mid 1960s until the mid 1980s, the profession seemed to lose interest in growth theory. Amartya Sen (1970, p9) in his introduction to a book of readings on Growth Economics says "With this intensely practical motivation it would have been natural for growth theory to take a fairly practice-oriented shape. This, however, has not happened and much of modern growth theory is concerned with rather esoteric issues. Its link with public policy is often very remote. It is as if a poor man collected money for his food and blew it all on alcohol."[3]

One factor contributing to the loss of interest was the results of Sato (1963). He noted that one can derive the speed of convergence from the Solow model, considered how long it would take in the neo-classical growth model for the economy to return to

[3]Sen thanks Frank Hahn, Robin Mathews, Luigi Pasinetti, Joan Robinson and Robert Solow for comments on an earlier draft of the introduction, saying "Their suggestions, often contradictory but always useful, have helped me a great deal to prepare the final version."

the steady state after some disturbance and showed that it may take a hundred years to cover 90 per cent, of the disturbance. Hahn and Matthews (1964) concluded that, insofar as Sato's finding is generally valid, steady-state solutions are likely to be of very limited value as an approximation to reality. While others questioned the algebra of slow return to steady state,[4] this pessimistic conclusion was influential.

Interest in growth theory revived in the mid 1980s, with theoretical papers like Romer (1986). By then, there was a lot more data available, in particular that resulting from the work by Maddison (1982) and Summers and Heston (1988). This prompted attempts to estimate the extent of convergence. An early attempt was Baumol (1986), who showed convergence among the industrialised market economies over the long period 1870-1979. This was criticised by DeLong (1988) on two counts. First, it suffered from sample selection bias, the sample consisted of those who had converged over the period and excluded once rich countries, like Argentina, who had not converged. Secondly, there was a problem of measurement error, which was much larger in the 1870 data. As discussed below, this biased the estimate of β to suggest more convergence than was the case. Baumol and Wolff (1988) responded and largely accepted the specific criticisms. However, the procedure Baumol used was widely adopted, though applied to different samples. The procedure involved estimating a cross country, cross section, auto-regression of the growth rate in income, the change in log GDP per capita, over a period on the initial value of log GDP per capita. The coefficient of the initial value provided what was subsequently labelled a measure of absolute or unconditional beta convergence.

Within growth theory the distinction between level and growth effects is important, but the cross section convergence regression can be written with either the growth rate, $\Delta y_{it} = y_{it} - y_{i,t-1}$, or the level of log per capita GDP, y_{it}, as dependent variable. For a suitable time interval, with $t = 1$ is 1979, $t = 0$ is 1870 equation (1) can be written with $y_{i1} = y_{it}$, $y_{i0} = y_{,it-1}$, $\beta y^*_{it} = \alpha$, $\beta_{it} = \beta$. It can be written in growth rate or level as dependent variable, in the case of the level version the coefficient of initial income

[4]This included a different Sato (1966).

is $\rho = 1 - \beta$:

$$\Delta y_{i1} = \alpha - \beta y_{i0} + u_i \tag{3}$$
$$y_{i1} = \alpha + \rho y_{i0} + u_i, \tag{4}$$

Using (4) we can examine the relationship between beta and sigma convergence. Express the data in terms of deviations from the cross section averages: $\tilde{y}_{i1} = (y_{i1} - \bar{y}_1)$, $\bar{y}_1 = N^{-1} \sum_{i=1}^{N} y_{i1}$, $\tilde{y}_{i0} = (y_{i0} - \bar{y}_0)$, $\bar{y}_0 = N^{-1} \sum_{i=1}^{N} y_{i0}$, $\tilde{u}_i = (u_i - \bar{u})$, $\bar{u} = N^{-1} \sum_{i=1}^{N} u_i$. Taking the mean squared deviation from the means in (4) gives an expression for the evolution of the variance,

$$N^{-1} \sum_{i=1}^{N} \tilde{y}_{i1}^2 = \rho N^{-1} \sum_{i=1}^{N} \tilde{y}_{i0}^2 + N^{-1} \sum_{i=1}^{N} \tilde{u}_i^2,$$

$$\sigma_1^2 = \rho \sigma_0^2 + \sigma_u^2.$$

Thus the variance of log per capita output is declining if

$$\frac{\sigma_1^2}{\sigma_0^2} < 1, \;\; \rho < 1 - \frac{\sigma_u^2}{\sigma_0^2}. \tag{5}$$

Since $\sigma_u^2 / \sigma_0^2 > 0$, then $\rho < 1$ is a necessary, but not a sufficient condition for the variance to decrease. The evolution of the variance also depends on the variance of the shocks, σ_u^2. Notice the final term in (5) is close to the R^2 of the regression $R^2 = 1 - \sigma_u^2 / \sigma_1^2$, so the worse the fit of this regression the further below one ρ needs to be (the larger β) for the variance to decline. This issue was noted both by Friedman (1992) and by Quah (1993), who also provides a dynamic model. It is called the Galton fallacy, since Francis Galton, the eugenicist who developed regression, had estimated an equation explaining the height of children by the height of their parents and noted that $\rho < 1$, which he called regression to the mean, the origin of the term regression.

As noted above, 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 output 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. Lee, Pesaran and

Smith (1997) consider other paths for the variance produced by more complex time series processes.

To appreciate the point about measurement error made by DeLong (1988), use deviations from the mean as defined above. Then current income is a function of true initial income: $\tilde{y}_{i1} = \rho\tilde{y}^*_{i0} + \varepsilon_i$, and measured initial income $\tilde{y}_{i0} = \tilde{y}^*_{i0} + \eta_i$, is true initial income $\tilde{y}^*_{i0}$, plus a measurement error η_i. Then under the classical measurement error assumptions and using $V(.)$ to denote variances and $C(..)$ to denote covariances, $V(\tilde{y}^*_{i0}) = \sigma^2_*$; $V(\varepsilon_i) = \sigma^2_\varepsilon$; $V(\eta_i) = \sigma^2_\eta$; $C(\tilde{y}^*_{i0}, \eta_i) = C(\tilde{y}^*_{i0}, \varepsilon_i) = C(\eta_i, \varepsilon_i) = 0$.

For large samples, the observed moments are

$$S_{11} = N^{-1}\sum_{i=1}^{N}\tilde{y}^2_{i1} = \rho^2\sigma^2_* + \sigma^2_\varepsilon$$
$$S_{00} = N^{-1}\sum_{i=1}^{N}\tilde{y}^2_{i0} = \sigma^2_* + \sigma^2_\eta$$
$$S_{10} = N^{-1}\sum_{i=1}^{N}\tilde{y}_{i1}\tilde{y}_{i0} = \rho\sigma^2_*$$

The direct regression coefficient is

$$\hat{b}_1 = \frac{S_{10}}{S_{00}} = \frac{\rho\sigma^2_*}{\sigma^2_* + \sigma^2_\eta} \leq \rho$$

so $\hat{b}_1 = \rho$ if $\sigma^2_\eta = 0$; while the reciprocal of the reverse regression coefficient is

$$\hat{b}_2 = \frac{S_{11}}{S_{10}} = \frac{\rho^2\sigma^2_* + \sigma^2_\varepsilon}{\rho\sigma^2_*} \geq \rho$$

so $\hat{b}_2 = \rho$ if $\sigma^2_\varepsilon = 0$. Notice

$$\frac{\hat{b}_1}{\hat{b}_2} = \frac{S^2_{10}}{S_{00}S_{11}} = r^2.$$

Thus ρ is asymptotically set identified, $\hat{b}_2 \geq \rho \geq \hat{b}_1$, and the better the fit of the regression, the smaller the possible set. Since $\beta = 1 - \rho$, the direct regression estimate of ρ, $\hat{b}_1$ which is biased down, will give an estimate of β that is biased up, indicating faster convergence. The measurement error issue may not be so pressing if more recent data are used than the 1870 estimates used by Baumol.

4 Convergence conditional on observables

Since absolute convergence did not seem to hold for most countries, the cross-section model was extended by Barro (1991) and Mankiw, Romer and Weil, MRW, (1992) to include determinants of the steady state level of productivity. This provided a measure of conditional beta convergence. MRW's variables, which have been labelled enhanced Solow fundamentals, were the investment rate, population growth and human capital. They note "Thus, the Solow model does not predict convergence; it predicts only that income per capita in a given country converges to that country's steady-state value. In other words, the Solow model predicts convergence only after controlling for the determinants of the steady state, a phenomenon that might be called "conditional convergence." Thus (3) was extended to include a vector of determinants of steady state, $\mathbf{x}_t$:

$$\Delta y_{i1} = \alpha - \beta y_{i0} + \boldsymbol{\gamma}' \mathbf{x}_i + u_i \tag{6}$$

and assuming $\beta > 0$ each country converges to a different steady state:

$$y_i^* = \beta^{-1}(\alpha + \boldsymbol{\gamma}' \mathbf{x}_i)$$

Regressions of the form (6) came to be called Barro regressions. Barro and Sala-i-Martin (1992) found evidence of absolute convergence for US states and conditional convergence for countries, after holding constant a set of variables that proxy for differences in steady state characteristics. Dating $\mathbf{x}_i$ may raise the difficulty about whether they are predictors, dated at period 0, or determinants of the process, in which case they are dated at period 1. In steady state there would be no difference, so the issue would not arise.

There is, the further difficulty that there are a very large number of candidates for inclusion in $\mathbf{x}_i$. Sala-i-Martin (1997) is entitled "I just ran four million regressions". Many of these candidates are potentially endogenous: high expected growth prompts high rates of investment. In choosing between models there are then problems of multiple testing. Even if there is no relationship 5% of the candidates will be significant at the 5% level just by chance. Many of the potential candidates are highly correlated making distinguishing between them difficult. For instance, openness and country size are correlated, small countries have higher trade shares than big ones.

The estimates of the speed of conditional convergence in the literature tended to converge. Barro (2015) says "According to the 'iron law of convergence', countries eliminate gaps in levels of real per capita GDP at a rate around 2% per year. Convergence at a 2% rate implies that it takes 35 years for half of an initial gap to vanish and 115 years for 90% to disappear." This is very close to the Sato (1963) estimate, but interpreted as the answer to a different question. Although the formal calculations were the same, Sato asked: how long the economy would take to adjust back to steady state after a disturbance, like a change in policy? Barro asked: if a country was starting from an initial position far from steady state, how long it would take the economy to make the transition to steady state? Lucas (1988 p7) also noted the change in question from national growth to international economic development. "Both Solow and Denison were attempting to account for the main features of U.S. economic growth, not to provide a theory of economic development, and their work was directed at a very different set of observations from the cross-country comparisons I cited in my introduction." Durlauf (2009) quotes Solow expressing "cheerful skepticism" about cross section growth regressions.

5 Fixed effects in dynamic panels with homogeneous slopes

The initial cross section convergence studies were quickly followed by panel studies. Again for a suitable time interval, for instance decades, the estimated equation was of the familiar two way fixed effect, TWFE, form:

$$\Delta y_{it} = \alpha_i + \alpha_t - \beta y_{i,t-1} + \boldsymbol{\gamma}' \mathbf{x}_{it} + u_{it}. \tag{7}$$

This has slope homogeneity, since β and $\boldsymbol{\gamma}$ do not differ across countries. However, the time varying steady states for each country differ because α_i and $\mathbf{x}_{it}$ differ across countries. Assuming $\beta > 0$, the steady states are given by

$$y_{it}^* = \beta^{-1} \left(\alpha_i + \alpha_t + \boldsymbol{\gamma}' \mathbf{x}_{it} \right).$$

The $\mathbf{x}_{it}$ capture time varying observed country specific influences. The time fixed effects, α_t, capture global time varying unobserved factors, such as improvements in technology, that influence all countries equally. The country fixed effects, α_i, capture unobserved country specific factors that are constant over time. The $\mathbf{x}_{it}$ are potentially subject to control, though perhaps endogenous: only initially rich countries may be able to afford those values of $\mathbf{x}_{it}$. The α_i are only fixed in that particular sample and may change if one changes the window as KWY do.

Equation (7) can be written in levels with $\rho = 1 - \beta < 1$ as

$$y_{it} = \alpha_i + \alpha_t + \rho y_{i,t-1} + \boldsymbol{\gamma}'\mathbf{x}_{it} + u_{it}; \tag{8}$$

and solving recursively from initial states y_{i0}

$$y_{it} = \rho^t y_{i0} + \frac{1-\rho^t}{1-\rho}\alpha_i + \sum_{j=0}^{t} \rho^j \alpha_{t-j} + \sum_{j=0}^{t} \rho^j \boldsymbol{\gamma}'\mathbf{x}_{i,t-j} + \sum_{j=0}^{t} \rho^j u_{i,t-j}.$$

If the number of observations T is small or ρ is close to one, e.g. 0.98 as Barro suggests, the econometric treatment of the initial conditions, y_{i0}, is important. In time series econometrics it is usually assumed that T is large, so the initial conditions do not matter. Bernard & Durlauf (1996) point out "In time series tests, one assumes that the data are generated by economies near their limiting distributions and convergence is interpreted to mean that initial conditions have no (statistically significant) effect on the expected value of output differences. Consequently, a given approach is appropriate depending upon whether one regards the data as better characterized by transition or steady state dynamics." ... "As a result, cross-section tests appear to more naturally apply to transition data whereas time series tests appear to more naturally apply to data whose sample moments well approximate the properties of the limiting distribution of the economies under study."

Panels, of course, combine cross section and time series data and the estimation issue is the relative weight given to the between country variation relative to the within country variation. Cross section estimators just use the between variation, fixed effect, FE, estimators the within variation, pooled OLS gives them equal weight. In deriving the asymptotic properties of the estimators one may let $N \to \infty$ for fixed T, which is common in panel studies, let $T \to \infty$ for fixed N, which is common in time series, or let them both go to infinity with, perhaps, some restriction on their relative rates, like $N/T \to 0$.

To consider the properties of these dynamic fixed effect estimators, consider a simpler model, where $\alpha_t = 0$, $\boldsymbol{\gamma} = \mathbf{0}$, so (8) is a stationary autoregression with $\rho < 1$,

$$y_{it} = \alpha_i + \rho y_{i,t-1} + u_{it}. \tag{9}$$

and u_{it} uncorrelated across groups and time. The limit of the FE estimator of ρ as $N \to \infty$ for fixed T is

$$P \lim_{N \to \infty} (\widehat{\rho}_{FE} - \rho) = -\frac{1+\rho}{T} + O(T^{-2}).$$

The FE estimator of ρ is consistent as $T \to \infty$, for fixed N; but inconsistent as $N \to \infty$ for fixed T. This Nickel (1981) (initial condition) bias is the result of the fact that the lagged dependent variable bias arising from the initial conditions is not removed by increasing N. Because the FE estimator takes deviations from the means, this induces a correlation between the error and the lagged dependent variable and this bias is amplified with N. If both $N \to \infty$ and $T \to \infty$, then to ensure consistency of the least squares estimates T must grow sufficiently fast relative to N, so that $N/T \to 0$. A range of bias corrected estimators have been suggested, and Chen, Chernozhukov and Fernandez-Val (2019) apply debiased estimators to the Acemoglu et al. (2019) democracy and growth question.

The derivations of the fixed effect bias in the convergence literature, for instance Acemoglu and Molina (2021) and Barro (2015) assume, like that above that $\rho < 1$. This is a strong assumption. When $\rho = 1$, the first term of the bias is much larger of the order $-3/T$ for large N rather than $-(1+\rho)/T$, Phillips and Sul (2007).

5.1 The role of fixed effects

Whether to include country fixed effects in panel estimators has proved controversial. Acemoglu and Molina (2021), like Acemoglu et al. (2019), argue in favour of including them to control for unobserved determinants of GDP per-capita across countries. The lack of country fixed effects in convergence models will then bias convergence coefficients towards zero, ρ toward one, since α_i and $y_{i,t-1}$ are positively correlated. In (9) suppose that $\alpha_i = \alpha + \eta_i$ omitting the fixed effect gives

$$y_{it} = \alpha + \rho y_{i,t-1} + \{\eta_i + u_{it}\} .$$

since $y_{i,t-1} = \alpha + \eta_i + \rho y_{i,t-2} + u_{i,t-1}$, the covariance of $\{\eta_i + u_{it}\}$ and $y_{i,t-1}$ will contain a term $E(\eta_i^2)$, which is positive.

Barro (2015) argues against including fixed effects. Both Acemoglu and Barro have taken the same position in earlier work. The argument for including them is relatively simple. They tend to be significant and since they are correlated with the laged dependent variable as shown above, and possibly also with other regressors, omitting them causes bias. The argument for excluding them is more subtle and rests on the choice of the parameter of interest.

Barro (2015, p915) says: "Inclusion of country fixed effects also affects the estimated coefficients and, especially, standard errors of explanatory variables – X variables – other than lagged dependent variables. Coefficients on country variables that are constant (such as geographical features and colonial history) cannot be estimated at all and variables that have little within-country time variation cannot be estimated with precision. In effect, the inclusion of country fixed effects throws out much of the information in isolating the effects of X variables on growth rates or other variables." ... "The perspective changes in the context of panel data observed for over a century. In this setting, the econometric problems posed by the inclusion of country fixed effects are less serious."

Barro (2015) in the abstract says "In a country panel since 1960, the estimated annual convergence rate for GDP is 1.7%, conditional on time-varying explanatory variables. With country fixed effects, the estimated convergence rate is misleadingly high. With data starting in 1870, country fixed effects are reasonable and the estimated convergence rate is 2.6%. Combining the two estimates suggests conditional convergence close to the 'iron-law' rate of 2%. With post-1960 data, estimation without country fixed effects reveals positive effects of GDP and schooling on law and order and democracy – consistent with the modernisation hypothesis. With post-1870 data, estimation without or with country fixed effects indicates modernisation."

KWY also argue that the country fixed effects absorb exactly the variation relevant for studying convergence. They say "These results suggest an interpretation that is consistent with neoclassical growth models. Conditional convergence has held throughout the period. Absolute convergence did not hold initially, but, as human capital, policies, and institutions, have improved in poorer countries, the difference in institutions across countries has shrunk, and their explanatory power with respect to growth and convergence has declined. As a result, the world has converged to absolute convergence because absolute convergence has converged to conditional convergence."

6 Heterogeneous slopes

6.1 Unit root issues

Relaxing slope homogeneity, and neglecting unobserved factors for a moment, consider the case of a single country and assume β_{it} in (1) is constant over time:

$$\Delta y_{it} = \beta_i(y^*_{it} - y_{i,t-1}) + u_{it}. \tag{10}$$

suppose that we assume a deterministic steady state process, where technology in a country grows at a constant rate g_i, that is $y^*_{it} = a_i + g_i t$ then we get

$$\Delta y_{it} = \beta_i(a_i + g_i t - y_{i,t-1}) + u_{it}. \tag{11}$$

Testing $H_0 : \beta_i = 0$, no convergence, against $H_1 : \beta_i > 0$, convergence, is the problem of testing for a unit root, the null hypothesis $H_0 : \rho = 1$, against the alternative $H_1 : \rho < 1$. This problem, familar from time series statistics, is a remarkably difficult one. Many years ago, Christiano and Eichenbaum (1990) asked. "Unit roots in real GNP: Do we know, and do we care?" Their answer to both questions was no and there is still no agreement as to whether or not there is a unit root in US log GDP. It is striking that a question that time series econometricians cannot answer, can be answered so precisely by growth economists using a cross-section. Of course, as noted earlier, these may be different questions. If there is a unit root the series is said to be integrated of order one, $I(1)$, it needs to be differenced once to make it stationary. If $\beta_i < 0$ it is said to be $I(0)$.

In practice, (11) is augmented by lagged changes in Δy_{it} to ensure that the error does not suffer serial correlation and this unit root test (based on the t ratio of the coefficient of y_{t-1}) is known as the Augmented Dickey Fuller (ADF) test. It has a non-standard critical value which has been tabulated. The tests have low power, a low probability of rejecting the null when it is false. This is a particular problem if the series is stationary but very persistent and β_i is as close to zero as suggested by the Barro's $\beta = 0.02$. Moving average components with coefficients close to -1 also cause problems because they cannot be approximated by autoregressive terms. Tests of the null hypothesis of stationarity, like the KPSS test of Kwiatowski et al. (1992) involve estimating the long run variance of a partial sum series which requires long time series if the size of the test is to be controlled particularly when the variable is stationary but highly persistent.

There are difficult trade-offs, for instance too few augmentation lags in the ADF equation result in size distortions whereas too many lags produce the correct size at the expense of power. The tests are sensitive: to the treatment of the deterministic elements such as intercept, trend and seasonal dummies; to the treatment of serial correlation either parametrically as in the ADF test, adding lagged changes, or non-parametrically as in the Phillips and Perron (1988) and KPSS tests, which correct the standard errors; to choice of tuning parameters such as lag length for parametric estimators and bandwidth or window size for non-parametric estimators; and to the presence of structural breaks.

Steady state growth does not characterise most countries, there are booms and slumps, accelerations and reversals. The presence of such shifts in mean or trend will tend to lead to non-rejection of the unit root null, despite the process being stationary around the shifting mean or trend. Distinguishing structural changes, large infrequent permanent innovations, from unit roots, small frequent permanent innovations, is difficult. Unit root tests are designed to determine whether the order of integration is zero or one, but it could be a fraction between zero and one. These fractionally integrated processes show long memory features which bear similarities to processes with occasional breaks.

The power of the test depends on the data span (number of years) not on the number of observations: long span time-series (over a century) tend to look I(0), short ones (over decades) I(1). But long span data are more likely to show structural breaks. There may be non-linearities. The series may look I(1) because the underlying process is a random walk within a band, but returns to the band very rapidly if it strays beyond the band.

Lee, Pesaran and Smith (1997) show that a stochastic Solow model implies that the general model is ARMA(2,1) with trend, and found a MA root of -1 in 76 out of the 102 countries examined using data 1965-1989. However, in all but 8 of the cases a common factor in the MA and AR parts could be removed, leaving a stationary AR1 with trend. They estimate (11) by maximum likelihood on the 102 countries, using deviations from cross section means, $y_{it} - \bar{y}_t$, to allow for a global factor. This gave an estimate of the mean of $\hat{\beta}_i = 0.3$ with a standard error of 0.02. As the econometric theory suggests assuming homogeneity, $\beta_i = \beta$, $g_i = g$, pushes the estimate of β towards zero and $\hat{\beta} = 0.04$. However, despite the large and precisely estimated mean value of $\hat{\beta}_i$ the ADF test where the lag order was chosen by the Bayesian Information Criterion, BIC, only rejected the unit root null in 14 of the 102 countries. Phillips and Sul (2009) allow for a global factor by using the ratio of log GDP per capita to its mean, $y_{it}/\bar{y}_t$, rather than deviations from the mean. Global factors are discussed further below.

One may add additional variables,

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i t + \boldsymbol{\delta}_i' \mathbf{x}_{it} + u_{it}.$$

If the x_{it} are $I(0)$ this may increase the power of the test. If the $\mathbf{x}_{it}$ are $I(1)$ they may be cointegrated with y_{it} and if $\beta_i < 0$, converging to a long run relationship $y_{it}^* = \beta_i^{-1}(\alpha_i + \gamma_i t + \boldsymbol{\delta}_i' \mathbf{x}_{it})$, which may corespond to a steady state.

Pesaran (2007a) applies the unit root tests to pairwise differences between countries in log per-capita output series from the Penn World Tables. Over 1961-2000 the unit root hypothesis was rejected at most in the case of 370 out of 4851 possible output gap pairs, just around 7.6% close to the nominal significance level of 5% used for the test. However, significant evidence of growth convergence is found, a result which is reasonably robust to the choice of the sample period and country groupings. Non-convergence of log per capita outputs combined with growth convergence suggests that while common technological progress seems to have been diffusing reasonably widely across economies, there are nevertheless important country-specific factors that render output gaps highly persistent, such that we can not be sure that the probability for the output gaps to lie within a fixed range will be non-zero.

One might hope to increase the power of the unit root tests by using a panel. However, there are a range of questions. How much heterogeneity in β_i across units is allowed under the alternative? It is homogeneous under the unit root null, $\beta_i = 0$. How do you interpret the null and alternatives? Rejecting the hypothesis that they are all $I(1)$ does not imply that they are all $I(0)$, only that a proportion are not $I(1)$. How do you combine the tests for the individual units? The Im Pesaran and Shin (2003) test averages the t ratios, but there are alternatives. How do you control for cross-section dependence? First generation tests assumed independence across groups, but the tests were very sensitive to the failure of this assumption, so second generation tests allow for cross section dependence, CSD, in various ways. Pesaran (2007b) allows for it by including cross section averages. There are two types of CSD, strong has global effects, a factor influencing all countries, weak has local effects, for instance spatial CSD where the correlations are among a relatively small group of neighboring countries.

6.2 Heterogeneity bias

Consider the case of a single x_{it}, which might be, for instance, US log per-capita GDP, if that represented the steady state,

$$y_{it} = \alpha_i + \rho_i y_{it-1} + \delta_i x_{it} + u_{it}. \tag{12}$$

The parameters of (12) can be efficiently estimated by OLS for each group, though there is a small T downward bias in $\hat{\rho}_i$. Suppose $\rho_i = \rho + \eta_{1i}$, $\delta_i = \delta + \eta_{2i}$, where $E(\eta_{ji}) = 0$, $j = 1, 2$. Then if we assume homogeneous slopes, we get

$$y_{it} = \alpha_i + \rho y_{it-1} + \delta x_{it} + \{\eta_{1i} y_{it-1} + \eta_{2i} x_{it} + u_{it}\}. \tag{13}$$

In this case as pointed out in Pesaran and Smith (1995) the FE estimator, which imposes homogeneity, gives biased estimates of $\rho = E(\rho_i)$ and $\delta = E(\delta_i)$ even for large T. The composite disturbance will be serially correlated if x_{it} is serially correlated, as it usually is, and so will not be independent of the lagged dependent variable.

Suppose x_{it} is generated by an AR1 process:

$$x_{it} = \mu_i(1 - \phi) + \phi x_{it-1} + \varepsilon_{it}$$

where μ_i is the unconditional mean. Then as x_{it} tends towards being an I(1) variable, i.e. as $\phi \rightarrow 1$, the Probability Limits of the FE estimator (taken by first letting $T \rightarrow \infty$. then letting $N \rightarrow \infty$) are given by:

$$\underset{\phi \rightarrow 1}{P\lim} (\widehat{\delta}_{FE}) = 0; \qquad \underset{\phi \rightarrow 1}{P\lim} (\widehat{\rho}_{FE}) = 1$$

irrespective of the true values of δ and ρ. If ϕ is positive, the usual case, the heterogeneity bias in $\widehat{\rho}_{FE}$ is upwards, the opposite of the Nickel bias. As usual, these results are not valid for $\phi = 1$, the $I(1)$ case, where the composite error in (13) contains $\eta_{2i}x_{it}$, so the error is also $I(1)$ and it is a spurious regression.

6.3 Models with cross section dependence

Consider the heterogeneous model where there is CSD which comes from an unobserved common factor. With a single time series one cannot allow for such an unobserved factor but with a panel one can. Pesaran (2006) suggests the common correlated effect, CCE, estimator that proxies unobserved common factors by cross section averages, Chudik and Pesaran (2015) extend it to the dynamic case.

Suppose that log per capital GDP, y_{it}, in each country is adjusting towards a common unobserved factor, a stochastic trend, f_t :

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i f_t + u_{it} \tag{14}$$

Averaging over countries and noting that if β_i and y_{it} are uncorrelated, the last term below goes to zero as N gets large

$$\Delta \bar{y}_t = \bar{\alpha} + \bar{\beta}\bar{y}_{t-1} + \bar{\gamma} f_t + \left(\sum_{i=1}^{N} (\beta_i - \bar{\beta}) y_{i,t-1}/N + \bar{u}_{it} \right).$$

Thus, as long as $\bar{\gamma} \neq 0$, we can proxy the factor by the cross section averages:

$$f_t = \bar{\gamma}^{-1} \left(\Delta \bar{y}_t - \bar{\alpha} - \bar{\beta}\bar{y}_{t-1} \right).$$

Inserting this into (14) gives an estimating equation for heterogeneous convergence:

$$\Delta y_{it} = a_i - \beta_i y_{i,t-1} + c_i \Delta \bar{y}_t + d_i \bar{y}_{t-1} + \varepsilon_{it}. \tag{15}$$

The Pesaran (2007b) panel unit root test allowing for cross section dependence uses the average t statistic on β_i from (15). As before rejection of the null hypothesis $\beta_i = 0$, only implies that not all of the $\beta_i = 0$, not that they are all non-zero.

The static long-run relationship when $\Delta y_{it} = \Delta \bar{y}_t = 0$ is

$$y_{it}^* = \beta_i^{-1}(a_i + d_i \bar{y}_t).$$

This makes it clear that if $\bar{y}_t$ is $I(1)$, y_{it} will also be $I(1)$, have a unit root even though $\beta_i \neq 0$. This will arise if the common factor f_t is $I(1)$, in which case y_{it} and $\bar{y}_t$ cointegrate if $\beta_i \neq 0$.

One can have a homogeneous long run relationship but heterogeneous short run coefficients including different rates of convergence to the long run common factor as in

$$\Delta y_{it} = a_i + \beta_i(\theta \bar{y}_{t-1} - y_{i,t-1}) + c_i \Delta \bar{y}_t + \varepsilon_{it}.$$

This is the pooled mean group estimator of Pesaran, Shin and Smith (1999). With an $I(1)$ factor, some countries can cointegrate $\beta_i > 0$, while others do not, $\beta_i = 0$.

If $\theta = d_i/\beta_i = 1$ then (15) is

$$\Delta y_{it} = a_i + \beta_i(\bar{y}_{t-1} - y_{i,t-1}) + c_i \Delta \bar{y}_t + \varepsilon_{it}; \tag{16}$$

with long run relationship

$$y_{it}^* = \frac{a_i}{\beta_i} + \bar{y}_t$$

If $y_{it} = \log Y_{it}$, then in the long run, each country would be proportional to the average, $Y_{it} = A_i \bar{Y}_t$ where $A_i = \exp(-a_i/\beta_i)$. To get $Y_{it} = \bar{Y}_t$, full convergence, we require $a_i = 0$, since for convergence we need $\beta_i \neq 0$.

Consider a steady state where $\Delta y_{it} = \Delta \bar{y}_t = g$ then (16) gives

$$g = a_i + \beta_i(y - \bar{y}) + c_i g; \tag{17}$$

$$y_{it}^* = \beta_i^{-1}\left[g(1 - c_i) - a_i\right] + \bar{y}_t.$$

So $A_i = \exp\left(\beta_i^{-1}\left[g(1-c_i) - a_i\right]\right)$ and depends on the rate of growth. $A_i = 1$ if $g(1-c_i) = a_i$, the intercept exactly offsets the growth, to stop you always failing to catch up with a moving target.

6.4 Interactive fixed effects

The model in equation (14) can be extended to

$$\Delta y_{it} = \alpha_i + \delta_t + \beta y_{i,t-1} + \boldsymbol{\gamma}'\mathbf{x}_{it} + \boldsymbol{\eta}_i\mathbf{f}_t + u_{it}.$$

The term $\boldsymbol{\eta}_i\mathbf{f}_t$ is known as an interactive fixed effect and here β and $\boldsymbol{\gamma}$ are homogeneous across i. Whereas the factor δ_t influences all countries equally, the factor $\mathbf{f}_t$ has different effects on each country.

Hayakawa et al. (2023) consider a small T case and follow Acemoglou et al. (2019) in regressing log GDP per capita y_{it} measured over five-year intervals on $y_{i,t-1}$, log investment-output ratio, log total factor productivity (TFP), log trade share in GDP, log infant mortality, and a dichotomous democracy variable. The data set used covers N = 82 countries with T = 5 five-yearly periods spanning 1981-2005. The regressions include country and time effects but also interactive fixed effects. They estimate the number of factors $m = 2$. The case with $m = 0$ is the standard TWFE model. Allowing for interactive fixed effects changes the estimate of β the coefficient of lagged log per capita GDP from 0.583 with $m = 0$ to 0.246 with $m = 2$. Most of the other coefficients change also. Allowing for cross-section dependence, CSD, is important.

The Acemoglou et al. (2019) study is also replicated with a different estimator in Dube et al. (2023). In the micro-econometric literature the TWFE estimator that has been used to estimate treatment effects using the difference in difference, DinD, approach has been subject to a range of criticisms. They propose a local projections in DinD estimator They do not find that their approach changes the conclusions about the impact of democracy but do not report results for the coefficient of the lagged dependent variable.

De Visscher et al. (2020) specify a Cobb-Douglas production function that parameterizes unobserved total factor productivity as a global technology process interacted with country specific absorptive capacity that varies stochastically over time and use a CCE approach to proxy global technology.

Mountford (2023) uses two related models. First, in (7) rather than the intercept, α_i, being constant it can move slowly through time, following a random walk

$$\alpha_{it} = \alpha_{i,t-1} + v_{it}.$$

This allows one to estimate a time varying steady state which adjusts in response to unobserved influences. Second, there is a hierarchical model where, in addition, the β_i are not constant, but randomly distributed. The models are estimated by Bayesian methods on data for US states and on the PWT cross country data and in both cases he finds that the results support the "Poor Stay Poor" hypothesis. Countries and US states are converging to different balanced growth paths though the technique treats the paths as reflecting unobserved variables rather than identifying specific determinants.

7 Conclusion

This survey has illustrated the wide variety of methods used to investigate convergence across countries in per capita GDP and, as noted above, there are large literatures investigating related questions about the relationship between growth and size for firms and cities and other variables.

The general argument of this survey has been that rather than there being different ways to measure a single concept, convergence, the different measures are measuring different aspects of the growth process, which cannot be summarised in a single statistic like the speed of adjustment or rate of convergence. Thus it is important to be specific about the question being asked, in particular convergence to what, and to tailor the question to the nature of the available data, in particular to the dimensions of N and T. One can estimate relatively simple models with small data-sets, more complex models with more parameters require larger data-sets. There is the inevitable trade-off between bias and efficiency: adding more parameters reduces the possibility of bias but also increases the variance of the estimators. However, even with a large potential data-set, there are many choices. These choices will be a matter of judgement, dependent on context. One may not want to use all the available countries, but pool ones that are relatively homogeneous. How to cluster countries remains an open question. The original data are usually annual, but many authors then take 5 or 10 year averages to remove cyclical effects. Whether this is a good idea remains an open question. Even if you have a long data-set you may wish to use a shorter time-series to avoid wars or structural breaks. How to determine the optimal length again is an open question. As noted in the introduction, one gets different a different answer to the question of whether there has been unconditional convergence if one starts the analysis in the year 2000 as compared to starting in 1960. The large literatures on all these issues does not provide any definite conclusions.

The growth process is complex, heterogeneous and mysterious, and is not well captured by the simple theoretical models commonly used in the convergence literature. Prasad (2023, p15) captures some of the mystery. "China has found a way to get results—generating sustained growth over a long period, improving the living standards of its people, avoiding a financial crisis, and pulling its economy through a number of perilous periods for the world economy. It has done all of this without a well-functioning financial system, a strong institutional framework, a market-oriented economy, or a democratic and open system of government. There is certainly cause for humility for anyone attempting to explain the China phenomenon based on the historical record and experiences of other countries. China's growth model and approach to reforms have not hewed to conventional norms and arguably tensions are building up in the system, with a possibly explosive meltdown at some point. But so far the government has proved adept at navigating around such perils."

Just as China suddenly started growing rapidly in 1979, Japan suddenly stopped growing rapidly in 1989 and the same may happen to China. While one can learn from the large N statistical studies, typical of the convergence literature, the range of special national characteristics means that the quantitative analysis needs to be supplemented by more qualitative historical studies. The challenges involved in combining the two sorts of studies are discussed in Smith (2021).

Acknowledgements. I am grateful to Kian Howe Ong for his encouragement in writing this paper and for useful comments from participants at a Nottinghan Ningbo workshop in September 2023 and from an anonymous referee.

Data availability statement. No datasets were generated or analysed during the current study.

Conflict of interest statement. The author declares he has no conflicts of interest.

References

Acemoglu, D., C.A. Molina (2021) Converging to converge? a comment. NBER working paper 28992.

Acemoglu, D., S. Naidu, P. Restrepo, J.A. Robinson, (2019). Democracy does cause growth. *Journal of Political Economy*, 127, 47-100.

Barro, R. J. (199I). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106, 407-44

Barro, R. J. (2015) Convergence and modernisation, *The Economic Journal* 125, 911-942.

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

Baumol, W. J., (1986) Productivity growth, convergence, and welfare: what the long-run data show, *American Economic Review*, 76, 1072-1085.

Baumol, W.J., E.N. Wolff (1988) Productivity growth, convergence, and welfare: Reply, *American Economic Review*, 78, 1155-1159.

Bernard, A.B., S. Durlauf (1995) Convergence in international output, *Journal of Applied Econometrics*, 10, 97-108.

Bernard, A.B., S. Durlauf (1996) Interpreting tests of the convergence hypothesis, *Journal of Econometrics*, 71, 161-73.

Bianchi, M. (1997) testing for convergence: evidence from non-parametric multimodality tests, *Journal of Applied Econometrics*, 12, 393-409.

Chen, S., V. Chernozhukov, I. Fernández-Val. (2019). Mastering panel metrics: causal impact of democracy on growth. *AEA Papers and Proceedings*, 109: 77-82.

Christiano, L.J., M. Eichenbaum (1990) Unit roots in real GNP: do we know and do we care? *Carnegie Rochester Conference Series on Public Policy* 31, 7-62.

Chudik, A., M.H. Pesaran (2015) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors, *Journal of Econometrics*, 188, 393-420.

DeLong, J. B., (1988) Productivity growth, convergence, and welfare: comment, *American Economic Review*, 78, 1138-54.

De Visscher, S., M. Eberhardt, G. Everaert (2020) Estimating and testing the multicountry endogenous growth model *Journal of International Economics* 125, 1-13.

Dube, A., D. Girardi, Ò. Jordà, A. M. Taylor (2023) A local projections approach to difference-in-differences event studies, NBER Working Paper No. 31184.

Durlauf, S. (2009) The rise and fall of cross country growth regressions, *History of Political Economy*, 41 (annual supplement) 315-333.

Fernández-Villaverde, J., L. E. Ohanian, W. Yao (2023) The neoclassical growth of China, NBER Working Paper 31351.

Friedman M. (1992) Do old fallacies ever die? *Journal of Economic Literature*, 30, 2129-32.

Galor, O. (1996) Convergence? inferences from theoretical models, *The Economic Journal*, 106, 1056-1069.

Hahn, F. H., R. C. O. Matthews, (1964) The theory of economic growth: a survey, *Economic Journal*, 74, 779-902.

Harrod, R. F. (1939) An essay in dynamic theory, *Economic Journal*, 49, 14-33.

Hayakawa, K., M.H. Pesaran, L. V. Smith (2023) Short T dynamic panel data models with individual, time and interactive effects, *Journal of Applied Econometrics,* forthcoming.

Im, K.S., M.H. Pesaran, Y. Shin (2003) Testing for unit roots in heterogeneous panels, *Journal of Econometrics*, 115, 53-74.

Johnson, P., C. Papageorgiou (2020) What remains of cross-country convergence, *Journal of Economic Literature,* 58, 129-175.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, 54, 159-178.

Kremer, M., J. Willis, Y. You (2022) Converging to convergence, *NBER Macroeconomics Annual*, 36, 337-412.

Lee, K., M. H. Pesaran, R. P. Smith (1997) Growth and convergence in a multicountry empirical stochastic Solow model, *Journal of Applied Econometrics*, 12, 357-392.

Lucas, R.E. (1988) On the mehanics of economic development, *Journal of Monetary Economics*, 22, 3-42.

Maddison, A., (1982) *Phases of Capitalist Development*, New York: Oxford University Press.

Mankiw, N.G., D. Romer, D. N. Weil, (1992). A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, 107, 407–37.

Mountford, A. (2023) Economic growth analysis when balanced growth paths may be time varying, MPRA 119938.

Müller, U.K., J.H. Stock, M.W. Watson (2022) An econometric model of international growth dynamics for long-horizon forecasting, *Review of Economics and Statistics*, 104, 857-876.

Nickel S (1981) Biases in dynamic models with fixed effects, *Econometrica*, 49, 1417-1426.

Patel, D., J. Sandefur, A. Subramanian (2021) The new era of unconditional convergence, *Journal of Development Economics*, 152, 102687.

Pesaran, M.H. (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* 74, 967-1012.

Pesaran, M.H. (2007a) A pair-wise approach to testing for output and growth convergence, *Journal of Econometrics*, 138, 312-355.

Pesaran, M.H. (2007b) A Simple panel unit root test in the presence of cross section dependence, *Journal of Applied Econometrics*, 22, 265-312.

Pesaran, M.H., Y. Shin, R.P. Smith (1999) Pooled mean group estimation of dynamic heterogeneous panels, *Journal of the American Statistical Association*, 94, 621-634.

Pesaran, M.H., R.P. Smith (1995) Estimation of long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics, 68, 79-114.*

Pesaran, M.H., R.P. Smith, T. Yamataga, L. Hvozdyk (2009) Pairwise Tests of Purchasing Power Parity, *Econometric Reviews*, 2009, 28, 1-27.

Phillips, P.C.B., P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Phillips, P.C.B., D. Sul (2007) Bias in dynamic panel estimation with fixed effects incidental trends and cross-section dependence, *Journal of Econometrics*, 137, 162-188.

Phillips, P. C. B., D. Sul (2009) Economic transition and growth, *Journal of Applied Econometrics* 24, 1153–85.

Prasad, E.S. (2023) Has China's growth gone from miracle to malady? NBER Working Paper 31151.

Quah, D. (1993) Galton's fallacy and tests of the convergence hypothesis, *Scandinavian Journal of Economics*, 95, 427-43.

Quah, D (1996) Empirics for economic growth and convergence, *European Economic Review* 40, 1353-75.

Rodrik, D. (2013). Unconditional convergence in manufacturing, *Quarterly Journal of Economics*, 128, 165–204.

Romer, P.M. (1986). Increasing returns and long-run growth, *Journal of Political Economy*, 94, 1002–37.

Sala-i-Martin, X. (1997) I just ran four million regressions, NBER Working Paper 6252.

Sato, K. (1966) On the adjustment time in neo-classical growth models. *Review of Economic Studies* 33, 263-68.

Sato, R. (1963) Fiscal policy in a neo-classical growth model: an analysis of time required for equilibrating adjustment, *Review of Economic Studies*, 30, 16-23.

Sen, A. (editor) (1970) *Growth Economics*, Penguin modern economics readings, Penguin Books, Harmondsworth.

Smith, R.P. (2021) The Challenge of identification and the value of descriptive evidence, Chapter 35, p917-940, of *The Palgrave Handbook of Comparative Economics*, Edited by Elodie Douarin and Oleh Havrylyshyn, Palgrave Macmillan.

Solow, R. M., (1956) A contribution to the theory of economic growth, *Quarterly Journal of Economics*, 70, 65-94.

Summers, R., A. Heston, (1988) A new set of international comparisons of real product and price levels: estimates for 130 countries, *Review of Income and Wealth*, 34, 1-25.

Sutton, J.,(1997) Gibrat's Legacy. *Journal of Economic Literature*, 35, 40-59.

Swan, T.W. (1956) Economic growth and capital accumulation, *Economic Record*, 32, 334-361.

Weill, L (2013) Bank competition in the EU: How has it evolved? *Journal of*

International Financial Markets, Institutions and Money, 25, 100-112.

Westerlund, J., H. Karabiyik, P.K. Narayan, S. Narayan (2022) Estimating the speed of adjustment of Leverage in the Presence of Interactive Effects, *Journal of Financial Econometrics*, 20, 942–960