EVALUATING TESTS FOR CONVERGENCE OF ECONOMIC SERIES USING MONTE CARLO METHODS WITH AN APPLICATION TO REAL GDP'S PER HEAD

Miguel Pedro Brito St. Aubyn

Thesis submitted for the degree of PhD in Economics
at the University of London

London Business School
August 1995

Abstract

The convergence concept can be found in different fields of the Economics discipline. Convergence of incomes or of GDPs per head is an important issue in growth theory. Some models of international trade imply the convergence of factor prices across countries. The European Union treaty includes some convergence rules on interest, exchange and inflation rates, and on budget deficits, for countries to enter a monetary union. This thesis starts by reviewing this literature.

These developments are accompanied by a number of empirical studies concerning the issue of measuring and testing for convergence. Proposed methods and provided results are in apparent contradiction. They are critically surveyed in Chapter 2.

The aforementioned disparate results constitute one of the main motivations for the systematic evaluation of the different methods. This is done resorting to simulation techniques using artificial data. Different techniques are assessed considering a number of different patterns of convergence. Chapters 3 to 7 include several experiments considering unconditional and conditional convergence, convergence clubs, limited and time-varying convergence as the true data generation process. Their results allow a better understanding of previous empirical studies and of the comparative advantages and weaknesses of different tests. In general terms, cross-sectional methods are not very reliable in the presence of cross-sectional heterogeneity. Time series methods do not share this disadvantage. A Kalman filter method is more robust to a time-varying speed of convergence when compared to cointegration techniques.

The thesis includes an empirical investigation on the convergence of GDPs per head across 16 industrialised countries using annual data from 1890 to 1989 (Chapter 8). The main conclusion is that countries tended to converge conditionally towards the US level, specially after the Second World War, at different speeds and to steady-state levels that are different from the pre-war values.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisor, Professor Stephen Hall, for his always invaluable and prompt advice.

I acknowledge financial support from JNICT (Junta Nacional de Investigação Científica), Portugal, and The British Council.

Introduction

Convergence is an important issue in Economics. There is a convergence debate going on in growth theory. In international trade theory, it is possible to find the terms "convergence of prices" applied either to goods or factors. The word "convergence" is also found in the European Union economic and political jargon. This thesis is more focused on the growth theory debate, but its results are meant to be relevant from a more general point of view.

By convergence of incomes per head in growth theory it is sometimes meant a tendency for this or other similar variable to become more or less equal across different countries. In one important qualification, "convergence" is sometimes considered to be "conditional." In this case, GDPs or incomes per head do not converge to the same levels, but differences between countries become stationary, so that growth rates are the same in the long-run. This is a consequence of the neoclassical growth model, and some attempts have been made to directly estimate it. Also, some technological catch up models allow for the existence of a limited convergence outcome: only countries with the necessary social capability converge to the income path of a leader country.

In fact, different models in growth theory have different implications for convergence of income levels: endogenous growth models suggest a tendency for divergence of incomes. Growth rates usually depend on country specific parameters or policies.

The recent surge of endogenous growth models in growth theory that challenge the convergence properties of the neoclassical model was accompanied by an increasing number of empirical studies that directly address the question of convergence of incomes across different economies (being them countries, states or regions within a country.) These studies propose different methods to test and measure convergence, and results using the same or similar data sets are apparently in contradiction when different approaches are considered.

At the same time, empirical studies in other areas of Economics have dealt with the convergence testing and measurement issue. In international trade theory, the empirical validation of the "factor price equalisation theorem" has lead some researchers into testing for convergence of time series of prices of factors. The Maastricht treaty on European Union and the convergence requirements contemplated in it have raised the concern in testing for convergence of interest, exchange and inflation rates, and of budget deficits as percentage of the GDP across European countries.

The techniques used in different strands of the literature are not always the same, but due to the similarity of concepts involved, some of them can be considered as adaptable and may be used in different contexts.

The fact that results using different techniques applied to the same data set are in apparent contradiction was one of the main motivations for this thesis. It was felt that a deeper understanding of the properties of different methods of testing and measuring convergence was needed so that existing results could be correctly interpreted and new results could be obtained with more confidence.

No real data is used to evaluate the different methods. In its stead, their assessment is made by employing simulation techniques and different Monte Carlo studies are constructed to assess the various methods of testing for convergence. A typical experiment consists of the following steps:

1 - To generate several replications with several artificial series converging according to a pre-specified pattern of convergence (e. g. complete unconditional convergence or different convergence clubs.)

2 - To apply different convergence tests and methods of measuring convergence to each replication and compute a number of statistics of interest (e. g. the number of times the "no convergence" null hypothesis was rejected.)

3 - To compare the performance of different methods under a similar convergence situation.

All the programmes used in these simulations were written on Gauss by the author and are available on request. A number of routines are included in an appendix to this thesis.

Once the properties of the different techniques were better understood, an empirical research using real data was done, both to illustrate some of the points that arose from the experiments and to give a contribution to a the ongoing debate on convergence of GDPs per head.

The first chapter in this thesis ("The Importance of Convergence in Economics") starts with a definition of convergence that tries to encompass the often only implicit definitions that can be found in the literature. It then surveys the theoretical literature in Economics that deals with convergence, with an emphasis on growth theory, since most of the empirical results and methods resulted from growth studies and were designed to deal with the convergence of income levels across economies. Sections on the international convergence of factor prices and on nominal convergence in the European Economic and Monetary Union are also included.

The second chapter ("Measuring Convergence: Methods and Results") reviews the main techniques and their outcomes when applied to measuring convergence of real data sets. A first appraisal of the methods is made here, and some apparent contradictions in results (or "puzzles") are highlighted.

Chapter 3 to 7 present the different experiments and provide an evaluation of the different techniques under different types of convergence.

In Chapter 3 ("Unconditional Convergence"), the data generation process (DGP, for short) is "well behaved", in the sense that all the series converge at the same speed to the same leader series. In the long run all the series tend to the same level, except for a stationary disturbance. The DGPs in the following chapters are departures from this "good behaviour." This chapter also includes a presentation of the methods as they are considered in the experiments to come.

In Chapter 4 ("Conditional Convergence") the series converge to the same leader, but their long run difference is not necessarily zero. It is still true, though, that their long run growth rate is the same. This kind of framework is compatible, in growth theory, with the theoretical results derived from the neoclassical growth model. This simple departure from unconditional convergence already puts some strains on the performance of some of the considered techniques.

The DGP in Chapter 5 ("Convergence Clubs") imposes the existence of two groups of series that converge to two different leaders. This hypothesis is more or less explicitly dealt with by some of the techniques and these are assessed here.

Chapter 6 ("Limited Convergence") includes experiments where only part of the series converges to one leader. The other series do not converge at all. This is the theoretical outcome of some growth models like the ones based on the "technological catch up" idea.

Chapter 7 ("Time-varying Convergence") considers a situation where the different series only start to converge some periods after the "initial period." In practice, the researcher would like to consider the hypothesis that convergence starts to occur somewhere after the beginning of the available time series, without being able to exactly locate that period. An understanding of the properties of the different techniques of testing and measuring convergence is therefore needed to face this situation.

After the evaluation of the different methods, an empirical study is presented in Chapter 8 ("Convergence Across Industrialised Countries (1890-1989).") Here, some new findings are introduced concerning the convergence of fifteen industrialised countries incomes per head to the United States level using a century of yearly data[1]. This study makes use of the results from the previous chapters.

The thesis ends with a general conclusion.

[1]The countries considered are the Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and the United States as a leader or benchmark country.

Chapter 1

The Importance of Convergence in Economics

Introduction: What is convergence of economic series?

The word convergence is often used with different meanings by different authors or in different papers. Quite often, there is no explicit definition of the idea. One has to infer the underlying definition from the whole text. Of course, there are many convergence definitions in the statistical and mathematical literature. These are in some sense connected to the issue herein studied, and as it will be shown shortly, this is a source of some confusion.

Firstly, different fields of economic theory and practice that deal with this idea are briefly referred. Then some rigorous definitions of convergence thought to be appropriate for economic series are given. These definitions are believed to underlie an important part of the literature on the subject.

In growth theory, "convergence" usually means a tendency for poorer countries to catch up with richer ones through time, so that economic conditions become eventually more similar among economies. Usually, a single series summarises these economic conditions. Most of the times, the series is GDP per head or GDP per active person. The first section of the first chapter surveys the developments on convergence in the growth theory literature.

In international trade theory, it is possible to find the terms "convergence of prices" applied either to goods or factors. Usually, "convergence" is considered different from "full equalisation."[1] "Convergence of international prices" would be observed if there is a tendency to a tightening of their distribution as measured by some dispersion measure. This chapter's second section deals with convergence of international factor prices.

The word "convergence" is also found in the European Union economic and political jargon. While the term "real convergence" refers to the growth theory meaning, "nominal

---

[1]As in Tovias (1982), for example.

convergence" covers convergence in interest, inflation and exchange rates, and in budget deficits as percentage of GDP. This kind of convergence has a juridic expression in the Maastricht treaty. The third section in this chapter handles with this kind of issues.

The aim is to encompass all these meanings of convergence in a more general setting[2], so that measuring methods can have a sense without referring to specific series.

Consider two economic series X_t and Y_t. These two series converge if:

$$(X_t - Y_t) \xrightarrow{P} \epsilon_t \ as\ t \to \infty, \tag{1.1}$$

where ϵ_t is a random variable obeying the following conditions:

$$E(\epsilon_t) = D_{xy}, \tag{1.2}$$

and

$$Var(\epsilon_t) = \sigma < \infty. \tag{1.3}$$

Equations (1.1) to (1.3) mean that the difference between the two series converges in probability to a third series that is stationary, having a constant mean D_{xy} and a constant variance σ.

It should be clear that although X_t converges to Y_t in economic terms, it does not converge in statistic terms. Nevertheless, a statistical meaning of convergence is used in the economic definition of the same term.

For reasons that will become apparent later, economic convergence is:

a) *point wise*, if $Var(\epsilon_t) = 0$;

b) *unconditional,* if $D_{xy} = 0$;

[2] This follows an idea taken from Hall, Robertson and Wickens (1993). Here, the convergence definitions are somewhat different.

c) *conditional*, if $D_{xy} \neq 0$.

The definitions above encompass the "beta-convergence" concept proposed by Barro and Sala-i-Martin (1992a, 1995). In their work, series are assumed to converge to their steady-state level at an annual constant rate. If the steady-states are the same, "beta-convergence" is unconditional. If they are different but grow at the same rate, "beta-convergence" is conditional. The definitions presented earlier in this introduction are more general because they do not imply a constant rate of convergence and therefore differences between series are not necessarily stationary from the beginning.

One possible formalisation of series that converge at a constant rate follows. It is one of the simplest forms of convergence but, with minor variations, it can be found in different theoretical models that predict convergence. It is also a good starting point for further extensions and it serves as an illustration for the different concepts of convergence previously defined.

Consider an attracting series X_1 and n-1 attracted series denominated by X_i, with i varying from 2 to n. The first series is a random walk with a drift:

$$x_{1,t} = x_{1,t-1} + g + \epsilon_{1,t}, \tag{1.4}$$

where t is a time subscript.

The attracted series are generated according to:

$$x_{i,t} = x_{i,t-1} + g + \beta(x_{1,t-1} - x_{i,t-1} - \overline{d}_i) + \epsilon_{i,t}, \tag{1.5}$$

so that they include an error correction term. In both equations ϵ is a white noise random variable, uncorrelated to previous values of x. β is the speed of convergence and is comprised between 0 and 1. $\overline{d}_i$ is a series-specific constant.

The difference between the attractor and any attracted series is therefore given by:

$$x_{1,t} - x_{i,t} = d_{i,t} = (1-\beta)d_{i,t-1} + \beta\overline{d}_i + \eta_{i,t}, \tag{1.6}$$

where η is the difference between the ϵs. In the long run (as t tends to infinity) the difference between the series becomes:

$$d_{i,\infty} = \overline{d}_i + \upsilon, \tag{1.7}$$

where υ depends on the ηs.

It can be inferred from equation (1.7) that series i converges to series 1:

- conditionally, if $\overline{d}_i$ is different from zero. In this case, the long run difference between the two series is a stationary random variable with a mean different from zero;

- unconditionally, if $\overline{d}_i$ equals zero;

- point wisely, if the variance of the ϵs, and consequently of η, is zero.

For some authors, convergence means a tendency for the cross-section dispersion of series to diminish over time. This concept is called "sigma-convergence" by Barro and Sala-i-Martin (1992a, 1995)[3]. It is shown below, using the same example, that convergence according to the earlier definition does not imply "sigma-convergence."

If the long run difference $\overline{d}_i$ is added to each attracted series, the resulting "parallel" series converges unconditionally to series 1 and is given by:

$$x_{i,t} + \overline{d}_i = x^c_{i,t} = x^c_{i,t-1} + g + \beta(x_{1,t-1} - x^c_{i,t-1}) + \epsilon_{i,t}, \tag{1.8}$$

The cross-sectional variance of $x^c_{i,t}$ is equal to:

$$\sigma^2_{x^c_t} = (1-\beta)^2 \sigma^2_{x^c_{t-1}} + \sigma^2_{\epsilon}. \tag{1.9}$$

Noting from equation (1.8) that $x^c_{i,t}$ does not depend on $\overline{d}_i$ and recalling that:

$$x_{i,t} = x^c_{i,t} - \overline{d}_i, \tag{1.10}$$

[3]Other examples of the use of this definition are Tovias (1982), Baumol and Wolf (1988), Mokhtari and Rassekh (1989) and Lichtenberg (1994).

it results that:

$$\sigma^2_{x_t} = \sigma^2_{x_{c_t}} + \sigma^2_{\bar{d}}. \qquad \textbf{(1.11)}$$

Using (1.9) and (1.11), it can be concluded that:

$$\sigma^2_{x_t} = (1-\beta)^2 \sigma^2_{x_{t-1}} + \beta(2+\beta)\sigma^2_{\bar{d}} + \sigma^2_{\epsilon}. \qquad \textbf{(1.12)}$$

In the long run, the cross-section standard deviation is equal to:

$$\sigma_x = \sigma^2_{\bar{d}} + \frac{1}{\beta(2+\beta)} . \sigma^2_{\epsilon} \qquad \textbf{(1.13)}$$

The long run cross-section variance positively depends on the variance of the shocks that affect each economy and also on the variance of $\bar{d}$. This last variance reflects the variation in the different steady-state levels. If convergence is point wise and unconditional, both these variances are zero and the cross-section standard deviation becomes also zero, so that all series coincide precisely.

From equation (1.12), and taking into account that β is comprised between 0 and 1, it results that a declining or increasing time path for $\sigma^2_{x_t}$ are both compatible with any of the definitions of convergence (conditional or unconditional.) $\sigma^2_{x_t}$ declines through time if it starts from a value higher than its long-term mean, and conversely, will increase if the series start too close together. For example, in case of unconditional convergence, if the series start all at the same point, the cross-section variance is expected to increase, driven by the cumulative effect of the series-specific random shocks.

1. Growth theory and convergence

1.1 The neoclassical model without growth

Solow (1957) neoclassical growth model remains a useful reference for the understanding of more recent developments on growth theory and convergence. In its simplest version, the Solow model exhibits no long run growth. In the steady-state, income and (physical) capital per head remains constant. Most neoclassical or endogenous growth models surveyed in this review are better understood as departures from this simpler model. It is therefore important to identify the hypotheses that impede long run growth here, since at least one is usually dropped in the other models.

Production at a given time (Y) depends on capital (K) and labour (L)[4], A being a country specific constant:

$$Y = AF(K,L). \tag{1.14}$$

This function (sometimes called the "neoclassical production function") is supposed to be homogeneous of degree one, twice differentiable and to obey the following conditions:

$$F_Z(K,L)>0, \qquad Z=K,L, \tag{1.15}$$

$$F_{KL}(K,L)=F_{LK}(K,L)>0, \tag{1.16}$$

$$F_{KK}(K,L)<0 \textit{ and } F_{LL}(K,L)<0, \tag{1.17}$$

$$\lim_{Z\to 0} F_Z(K,L)=+\infty, \qquad Z=K,L. \tag{1.18}$$

$$\lim_{Z\to\infty} F_Z(K,L)=0, \qquad Z=K,L, \tag{1.19}$$

meaning that marginal productivity is positive but decreasing to zero, being infinitely high when the factor is used in very small quantities. Conditions (1.18) and (1.19) are called the Inada (1963) conditions.

Since the production function is homogeneous, production per unit of labour depends on capital per unit of labour[5]:

---

[4]Time indexes are dropped when they are not necessary.

[5]Small letters denote "per unit of labour" variables.

$$y = \frac{Y}{L} = AF(\frac{K}{L},1) = Af(k). \qquad \textbf{(1.20)}$$

If labour grows at an exogenous rate n and capital depreciates at a fixed rate δ, then the law of motion for capital is:

$$\dot{k} = sAf(k) - (n+\delta)k, \qquad \textbf{(1.21)}$$

where the savings rate s is constant[6]. A variable with a "." stands for its time derivative.

Figure 1.1 represents equation (1.21). In that graph, curve S depicts savings as a function of capital per head (sAf(k)) and line D pictures capital per head depreciation ((n+δ)k).

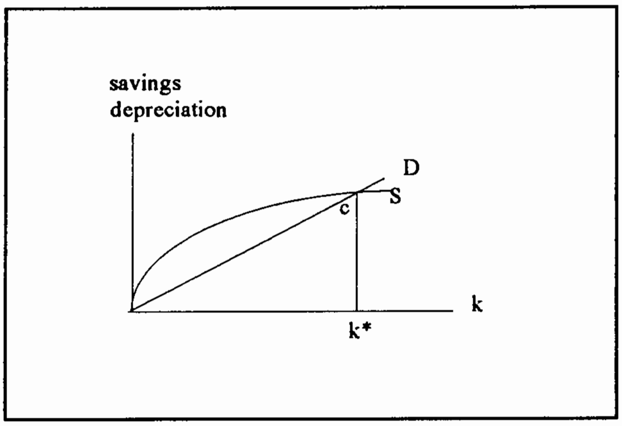

Figure 1.1
Savings and depreciation
in a zero-growth model

Since the marginal productivity of capital tends to zero, the slope of curve S is decreasing and zero in the limit[7]. Consequently, it crosses line D at some point c. Net investment is equal to the difference between S and D. To the left (right) of k^*, capital per head is increasing (decreasing) so that k^* is a stable equilibrium.

It can easily be concluded that this model only exhibits growth as part of its transitional

---

[6]The saving rate could be endogenously determined by some kind of optimising behaviour. The issue here is to make the hypotheses that impede growth explicit. As it will be apparent soon, these come from the production function specification and do not depend on consumers behaviour.

[7]The slope of S is equal to $sAf_k(k)$, $Af_k(k)$ being the marginal product of capital per head.

dynamics: capital per head and therefore production grows while they are lower than their long run (steady-state) value. Once that point is reached, there is no room for additional growth.

The following two hypotheses may be shown to be responsible for this result:

i) A is constant;

ii) the marginal productivity of capital tends to a value inferior to $(\delta + n)/s$.

Suppose that A increases through time. The marginal productivity of capital would not fall to zero, and capital per head could be accumulated indefinitely. In graphical terms, this would correspond to successive upward shifts in the S curve. This is the solution adopted by the exogenous growth neoclassical model, but also by several endogenous growth models.

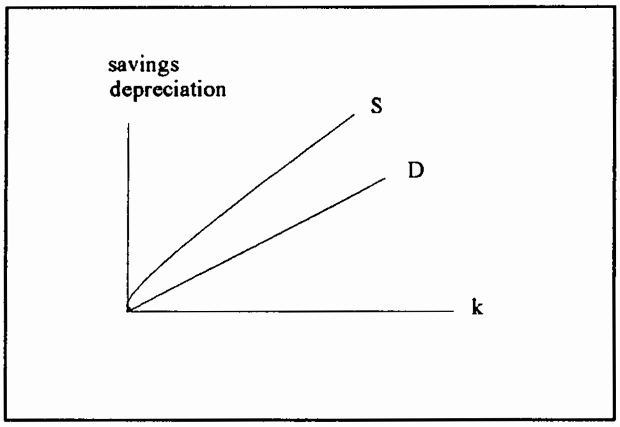

Figure 1.2
Savings and depreciation
in a positive growth model

Otherwise, consider that the marginal productivity of capital does not fall to zero, but tends to some positive quantity that is higher than $(\delta + n)/s$. Now, capital marginal product is asymptotically constant and positive and net investment is assured to remain positive. Net investment translates itself directly into growth. In graphical language, this means that curve S has a limiting slope that is higher than $\delta + n$, and never crosses line D. Figure 1.2 pictures this situation. As it will be shown later, a family of endogenous growth models rely on this kind of assumption.

1.2 The neoclassical model with exogenous growth

a) Optimising approach

As previously explained, growth can be introduced into the neoclassical model if it is assumed that A grows through time. This was Solow's approach in his 1957 paper. That kind of exogenous technical progress may be combined with optimising consumer behaviour, as done by Koopmans (1965).

Most writers use the same intertemporal utility function[8], maximised by a representative consumer, be their models exogenous or endogenous growth ones. Apparently, this revealed preference derives from the function's mathematical tractability. Some authors want to have their work compared with previous theoretical research, and this seems to contribute to the persistence of this use[9]. Some important insights would remain clear if, say, constant saving rates were assumed. This robustness of results is comfortable since it minimises the possible biases that could result from the arbitrary choice of a preferences pattern. Having written this, the tradition is followed and the neoclassical model is presented with exogenous technological progress and optimising behaviour.

The version that is presented below is the one used by Barro and Sala-i-Martin (1992a, 1992b, 1995) and also by King and Rebelo (1993) with minor differences. Chiang (1992) contains a succinct and clear mathematical explanation of the model.

The production function is the neoclassical one, but this time A_t is not constant and represents the labour augmenting technological progress, its growth rate being denoted by g:

$$Y_t = F(K_t,\ A_t L_t), \tag{1.22}$$

$$A_t = A_0 e^{gt}. \tag{1.23}$$

[8]Among others: Barro and Sala-i-Martin (1992a, 1992b, 1995), Chiang (1992), Helpman (1992), King and Rebelo (1990, 1993), Lucas (1988), Rebelo (1991, 1992) and Romer (1990a).

[9]As a matter of fact, this also makes the reviewer's task easier.

As before, labour is supposed to grow at the constant rate n:

$$L_t = L_0 e^{nt}. \tag{1.24}$$

It is useful to divide both terms of equation (1.22) by A_tL_t, to get a production function in "per unit of efficient labour" terms:

$$y = \frac{Y}{AL} = F(\frac{K}{AL}, 1) = f(k). \tag{1.25}$$

where small letters refer to per unit of efficient labour variables and time indexes are omitted.

The resource restriction may be expressed in per unit of efficient labour terms, as follows:

$$\dot{k} = f(k) - c - (g+n+\delta)k. \tag{1.26}$$

The representative consumer is supposed to maximise the following intertemporal utility function:

$$U = \int_o^\infty \frac{(C_t/L_t)^{1-\theta}}{1-\theta} L_0 e^{nt} e^{-\upsilon t} dt. \tag{1.27}$$

The momentary utility function exhibits constant marginal utility elasticity (θ). The representative consumer cares about consumption per head, weighted by population size. υ is a discount factor. It is possible to write the utility function in terms of per efficient unit of labour consumption (c_t). It can be shown that:

$$U = L_0 A_0^{1-\theta} \int_o^\infty \frac{c_t^{1-\theta}}{1-\theta} e^{(g+n-g\theta-\upsilon)t} dt. \tag{1.28}$$

To simplify, U is divided by $L_0A_0(1-\theta)$. The resulting utility function V is a monotonic transformation of U, and therefore represents the same preferences:

$$V = \int_0^\infty \frac{c_t^{1-\theta}}{1-\theta} e^{-\rho t} dt. \tag{1.29}$$

ρ is a discount rate that is assumed to be positive and that depends on other parameters:

$$\rho = \upsilon + g\theta - g - n \tag{1.30}$$

It is possible to show that a representative consumer that maximises V will choose a consumption path that obeys the following:

$$\frac{\dot{c}}{c} = \theta^{-1}(r - \rho), \tag{1.31}$$

where r is the real interest rate. Here, the relevant interest rate is equal to the net marginal productivity of capital:

$$r = f_k(k) - n - g - \delta. \tag{1.32}$$

Equations (1.26) and (1.31) describe the time path for consumption and capital and are therefore crucial to derive the steady-state and to describe the transitional dynamics.

When the economy is in the steady-state, capital and consumption per unit of effective labour remain constant. In algebraic terms, this means that the two following conditions hold:

$$\dot{c}_t = 0, \tag{1.33}$$

$$\dot{k}_t = 0, \tag{1.34}$$

From now on, the production function is assumed to be Cobb-Douglas, so that:

$$Y = K^\alpha (AL)^{1-\alpha}, \tag{1.35}$$

and

$$y = f(k) = k^{\alpha}. \tag{1.36}$$

The steady-state values for capital (k*) and consumption (c*) per unit of efficient labour are accordingly derived from equations (1.26) and (1.31), using (1.30), (1.32) and (1.36):

$$k^{*} = \left(\frac{g\theta+\delta+\upsilon}{\alpha}\right)^{\frac{1}{\alpha-1}}, \tag{1.37}$$

$$c^{*} = \left(\frac{g\theta+\delta+\upsilon}{\alpha}\right)^{\frac{\alpha}{\alpha-1}} - (g\theta+\upsilon)\left(\frac{\delta+\upsilon+(\theta-1)g}{\alpha}\right)^{\frac{1}{\alpha-1}}. \tag{1.38}$$

In the steady-state, consumption per capita and capital per capita grow at the same rate g, which happens to be the rate of technological progress.

Equations (1.26) and (1.31) are non-linear differential equations. One way to analyse consumption and capital behaviour close to the steady-state is to log-linearise these equations around it.

Applying logs to both equations results in:

$$\dot{\log c} = \frac{\dot{c}}{c} = \theta^{-1}(\alpha k^{\alpha-1} - (g+n+\delta+\rho)), \tag{1.39}$$

$$\dot{\log k} = \frac{\dot{k}}{k} = k^{\alpha-1} - \frac{c}{k} - (g+n+\delta). \tag{1.40}$$

Equations (1.39) and (1.40) can be rewritten as:

$$\dot{\log c} = \theta^{-1}(\alpha e^{(\alpha-1)\log k} - (g+n+\delta+\rho)), \tag{1.41}$$

$$\dot{\log k} = e^{(\alpha-1)\log k} - e^{\log c - \log k} - (g+n+\delta). \qquad \textbf{(1.42)}$$

In the steady-state, log c and log k remain constant. The following expressions are useful when deriving the results that follow them:

$$\dot{\log c} = 0 \Leftrightarrow e^{(\alpha-1)\log k^*} = \frac{g+n+\delta+\rho}{\alpha}, \qquad \textbf{(1.43)}$$

$$\dot{\log k} = 0 \Leftrightarrow e^{\log c^* - \log k^*} = \frac{(1-\alpha)(g+n+\delta)+\rho}{\alpha}. \qquad \textbf{(1.44)}$$

The derivatives that follow are evaluated at the steady-state and derived from (1.41) and (1.42), using (1.43) and (1.44):

$$\frac{\partial \dot{\log c}}{\partial \log c} = 0, \qquad \textbf{(1.45)}$$

$$\frac{\partial \dot{\log c}}{\partial \log k} = -\frac{(1-\alpha)(g+n+\delta+\rho)}{\theta}, \qquad \textbf{(1.46)}$$

$$\frac{\partial \dot{\log k}}{\partial \log c} = -\frac{(1-\alpha)(g+n+\delta)+\rho}{\alpha}, \qquad \textbf{(1.47)}$$

$$\frac{\partial \dot{\log k}}{\partial \log k} = \rho. \qquad \textbf{(1.48)}$$

The log-linearised system is as follows:

$$\begin{pmatrix} \dot{\log c} \\ \dot{\log k} \end{pmatrix} = \begin{pmatrix} 0 & -\mu_1 \\ -\mu_2 & \rho \end{pmatrix} \begin{pmatrix} \log c - \log c^* \\ \log k - \log k^* \end{pmatrix} \tag{1.49}$$

where:

$$\mu_1 = -\frac{\partial \dot{\log c_t}}{\partial \log k_t}, \tag{1.50}$$

and

$$\mu_2 = -\frac{\partial \dot{\log k_t}}{\partial \log c_t}. \tag{1.51}$$

The system matrix eigenvalues are λ_1 and λ_2:

$$\lambda_1 = \frac{\rho - \sqrt{\rho^2 + 4\mu_1\mu_2}}{2}, \tag{1.52}$$

$$\lambda_2 = \frac{\rho + \sqrt{\rho^2 + 4\mu_1\mu_2}}{2}. \tag{1.53}$$

Accordingly, the law of motion for capital is:

$$\log k_t - \log k^* = \eta_1 e^{\lambda_1 t} + \eta_2 e^{\lambda_2 t} \tag{1.54}$$

Noting that $\lambda_1 < 0$ but that $\lambda_2 > 0$, it is necessary to set $\eta_2 = 0$ to rule out any explosive behaviour that would violate the transversality conditions. The value for η_1 results from the initial condition:

$$\eta_1 = \log k_0 - \log k^*. \qquad \textbf{(1.55)}$$

The solution for log k_t is derived using expressions (1.54) and (1.55):

$$\log k_t = \log k^* + (\log k_0 - \log k^*)e^{\lambda_1 t}. \qquad \textbf{(1.56)}$$

Since λ_1 is negative, it is clear from equality (1.56) that k_t approaches k* asymptotically. Noting that $y_t = k_t^{\alpha}$, it follows that y increases (or decreases) at the same rate as k:

$$\log y_t = \log y^* + (\log y_0 - \log y^*)e^{\lambda_1 t}. \qquad \textbf{(1.57)}$$

Expression (1.57) can be transformed into the following:

$$\frac{\log y_t - \log y_0}{t} = -\frac{1-e^{\lambda_1 t}}{t}\log y_0 + \frac{1-e^{\lambda_1 t}}{t}\log y^*. \qquad \textbf{(1.58)}$$

This last equation shows that:

- the time average of income per unit of effective labour growth rate tends to zero over time;
- the time average of income per unit of effective labour growth rate is higher the smaller initial income is.

Recalling that:

$$y_t = \frac{Y_t}{(A_0 e^{gt})L_t} \Leftrightarrow \log y_t = \log\left(\frac{Y_t}{L_t}\right) - \log A_0 - gt, \qquad \textbf{(1.59)}$$

it is also possible to write an equation for the average growth rate of income per head, using (1.58) and (1.59):

$$\frac{\log\frac{Y_t}{L_t} - \log\frac{Y_0}{L_0}}{t} = g - \frac{1-e^{\lambda_1 t}}{t}\log\frac{Y_0}{L_0} + \frac{1-e^{\lambda_1 t}}{t}(\log y^* + \log A_0). \qquad \textbf{(1.60)}$$

The average growth rate of income per head approaches g, the rate of technical progress, as time tends to infinity. The average growth rate is the higher the lower initial income is.

Convergence properties of the neoclassical model

To assert the type of convergence that this model exhibits one can suppose that two countries share the same rate of exogenous technological progress g. After both have converged to the steady-state, both countries' income per head grow at the same rate g. This is already apparent from equation (1.60), but it is useful to derive an expression for income per capita once the economy reached the steady-state.

Income per head is given by the following equation, which is derived from the production function:

$$\frac{Y}{L} = \left(\frac{K}{L}\right)^{\alpha} A^{1-\alpha}. \qquad \textbf{(1.61)}$$

In the steady-state, capital per unit of labour is:

$$\frac{K}{L} = Ak^{*}. \qquad \textbf{(1.62)}$$

Substituting (1.62) into (1.61) it gives:

$$\frac{Y_t}{L_t} = A_t k^{*\alpha} = (A_0 k^{*\alpha}) e^{gt}. \qquad \textbf{(1.63)}$$

Let two countries be named A and B. Country A income per unit of labour relative to B is constant and equal to:

$$\frac{\frac{Y_a}{L_a}}{\frac{Y_b}{L_b}} = \frac{A_{o,a} k_a^{*\alpha}}{A_{o,b} k_b^{*\alpha}}. \qquad \textbf{(1.64)}$$

So equation (1.64) shows that even if two countries share the same production function, preferences and rate of technical progress (so that their steady-state capital per effective unit of labour is the same), they do not necessarily tend to the same income per capita, since they may well have different A_0s. Since their incomes will grow at the same rate, the logs of their incomes will differ by a constant.

Moreover, even if two countries share the same rate of technical progress, they tend to grow at the same rate g, so that the proportion between incomes per head stays constant, *even if all the other parameters are different.*

This means that, in general, the referred model assures *conditional point wise convergence* between (logs of) incomes per capita. The constant will be a function of initial conditions and preferences' parameters. It is of course possible to introduce some stochastic elements in the model in order to get a conditional convergence result that is not point wise.

b) A neoclassical model that includes human capital

Mankiw, Romer and Weil (1992) is an influential paper that, in its authors' words (p. 407), "takes Robert Solow seriously." It is an attempt to show that (p. 407) "an augmented Solow model that includes accumulation of human as well as physical capital provides an excellent description of the cross-country data."

Production level is supposed to depend on three factors: labour (L), physical capital (K) and human capital (H), the production function being Cobb-Douglas:

$$Y = K^{\alpha} H^{\beta} (AL)^{1-\alpha-\beta}, \qquad \textbf{(1.65)}$$

where, as before, A stands for the level of technology, progressing at a rate g, and L grows at a rate n.

This model does not include any utility maximization. Instead, investment in human and physical capital is constrained to be a fixed proportion of income, δ being the common depreciation rate:

$$\dot{h} = s_h y - (n+g+\delta) h, \qquad \textbf{(1.66)}$$

and

$$\dot{k} = s_k y - (n+g+\delta)k. \tag{1.67}$$

As before, small letters refer to quantities per effective unit of labour.

The implied levels of steady-state human and physical capital are:

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}}, \tag{1.68}$$

and

$$h^* = \left(\frac{s_k^{1-\alpha} s_h^{\alpha}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}}. \tag{1.69}$$

Approximating around the steady-state:

$$\frac{d\log y_t}{dt} = \lambda(\log y^* - \log y_t), \tag{1.70}$$

where λ (the *speed of convergence* to the steady-state) is given by:

$$\lambda = (n+g+\delta)(1-\alpha-\beta). \tag{1.71}$$

This result implies the following one:

$$\log y_t = \log y^* + (\log y_0 - \log y^*)e^{-\lambda t}. \tag{1.72}$$

The similarity between this last result and the one derived from the model without human capital is complete (see equation (1.57)), except for the concrete value of λ. However, Mankiw, Romer and Weil prefer to substitute for the value of y^* and derive the following "initial value" expression:

$$\log y_t - \log y_0 = (1-e^{-\lambda t})\frac{\alpha}{1-\alpha-\beta}\log s_k + (1-e^{-\lambda t})\frac{\beta}{1-\alpha-\beta}\log s_h - (1-e^{-\lambda t})\frac{\alpha+\beta}{1-\alpha-\beta}\log(n+g+\delta) - (1-e^{-\lambda t})\log y_0 \tag{1.73}$$

Growth of output per effective worker depends negatively on initial income, but the respective coefficient tends to 0 through time. It is worth noting that growth also depends on the saving rates and on the rates of population growth, technical progress and depreciation.

Following the same lines as done previously with the model without human capital, it can be shown that:

- when t tends to infinity, countries exhibit the same income per head growth rate;
- when transitional dynamics are still operating, per head income growth rate depends negatively on initial income and on the determinants of the steady-state, very much like the last equation shows.

It can therefore be concluded that both neoclassical models imply some sort of *conditional convergence* i.e. convergence of incomes per capita after conditioning on a set of variables that allow for steady-state differences.

1.3 Convergence and technological catch up: a non neoclassical approach

Several authors have given some explanations and developed theoretical arguments for levels of GDP per head convergence. These come either as a justification for an observed empirical pattern (especially among the OECD countries) or as a testable hypothesis.

An influential paper by Baumol (1986) provides empirical evidence in favour of convergence among OECD countries and of divergence of larger groups of countries, thus conducing to the idea that there is a "convergence club." His econometric methods will be surveyed later. Convergence is considered to result from the international public-good nature of successful productivity-enhancing measures. On the one hand, countries increasingly imitate innovations. On the other hand, investment also may exhibit international public good properties, even if the factor price equalisation theorem is not applicable.

Abramovitz (1986) provides the notion of "social capability": different social institutions and processes make some countries better or worse at catching up. Accordingly, some forge ahead while others fall behind.

There is an important difference between these explanations for convergence and the ones that come from neoclassical models. The latter usually assume the same rate of technological progress in every country (at least the ones that belong to the same club), convergence being explained by the accumulation of capital conditioned on possibly different steady-states. The reader can think of the former as models in which each country exhibits different rates of technical progress, the converging or catching up countries having a higher growth rate than the technological leader.

Dowrick and Gemmell (1991) emphasise that catching up effects may differ across sectors (agriculture and industry, for instance) and across countries. Moreover, differences in GDP growth rates may also result from different growth rates of factor inputs.

Dowrick and Nguyen (1989) formalise this last point. Country i output at time t is supposed to obey the following equation:

$$\log Y_{i,t} = A_i + \alpha \log K_{i,t} + \beta \log L_{i,t} + \gamma t + \lambda \log F_{i,t}, \qquad \textbf{(1.74)}$$

where γ is the leader's rate of technological progress (the rate of growth in "total factor productivity" in the leading country), λ is a positive parameter smaller than one and $F_{i,t}$ is a catch up function, defined as:

$$\frac{F_{i,t}}{F_{i,t-1}} = \frac{\frac{Y_{1,t-1}}{L_{1,t-1}}}{\frac{Y_{i,t-1}}{L_{i,t-1}}}, \qquad \textbf{(1.75)}$$

where country 1 is the leader. From these, Dowrick and Nguyen (1989) derive the following:

$$\log\frac{Y_{i,t}}{L_{i,t}} - \log\frac{Y_{1,t}}{L_{1,t}} = \alpha(k_i - k_1) + (\beta - 1)(l_i - l_1) + (1-\lambda)(\log\frac{Y_{i,t-1}}{L_{i,t-1}} - \log\frac{Y_{1,t-1}}{L_{1,t-1}}), \qquad \textbf{(1.76)}$$

where k_i and l_i stand respectively for the constant capital and labour growth rates in country i.

From this last equation results the following expression for the "final year" relative output per worker:

$$\log\frac{Y_{i,T}}{L_{i,T}} - \log\frac{Y_{1,T}}{L_{1,T}} = \frac{1-(1-\lambda)^T}{\lambda}[\alpha(k_i - k_1) + (\beta - 1)(l_i - l_1)] + (1-\lambda)^T(\log\frac{Y_{i,0}}{L_{i,0}} - \log\frac{Y_{1,0}}{L_{1,0}}). \qquad \textbf{(1.77)}$$

If one takes the limit of expression (1.77) as time tends to infinity, it results that:

$$\lim_{t\to\infty}\left(\log\frac{Y_{i,T}}{L_{i,T}}-\log\frac{Y_{1,T}}{L_{1,T}}\right)=\frac{1}{\lambda}[\alpha(k_i-k_1)+(\beta-1)(l_i-l_1)] \tag{1.78}$$

From expression (1.78), it can be concluded that the difference between the logs of incomes per capita tends to a constant. This constant depends on the relative growth rates of capital and labour.

The average growth rate of income per unit of labour can be shown to be equal to:

$$g_{Y/L_i}=c+\frac{\delta}{\lambda}[\alpha k_i+(1-\beta)l_i]-\delta[\log\frac{Y_{i,0}}{L_{i,0}}-\log\frac{Y_{1,0}}{L_{1,0}}], \tag{1.79}$$

where

$$\delta=\frac{1-(1-\lambda)^T}{T}, \tag{1.80}$$

and

$$c=\gamma+(1-\frac{\delta}{\lambda})(\alpha k_1+(\beta-1)l_1). \tag{1.81}$$

Since $\lambda>0$ the average growth rate of a country income per capita depends negatively on the initial income value. The coefficient approaches zero as time goes by. This is not without a parallel with the results from the neoclassical model. Also note that the average growth rate approaches the growth rate of the leading country as time tends to infinity. In fact, it happens that:

$$\lim_{T\to\infty} g_{Y/L_i}=\gamma+\alpha k_1+(\beta-1)l_1. \tag{1.82}$$

It can be concluded that two different countries end up growing at the same rate as that of the leading country, so that the logs of their incomes will differ by a constant. The model implies a form of *conditional convergence*, allowing for some transitional dynamics. Convergence is point wise, because there is no stochastic element. This result is very similar to the ones derived from the neoclassical models discussed before.

1.4 Endogenous growth models[10]

a) Growth and returns from capital

As previously discussed, a main obstacle to growth in the simple neoclassical model is the decreasing returns to capital *and* their tendency to a quantity that is sufficiently close to zero. If returns to capital are decreasing but do not tend to zero, it is possible to have endogenous growth (meaning that the growth rate depends from the model parameters values.) Jones and Manuelli (1990) model exploits this idea. The implied production function is of the type:

$$Y_t = AK_t + BK_t^{\alpha} L_t^{1-\alpha}. \quad \textbf{(1.83)}$$

Since it does not affect the relevant results, one can assume that labour is constant (make it equal to one) and that there is no depreciation.

Clearly returns to capital tend to A, supposed to be a constant. In the long term, the time change of capital is:

$$\dot{K} = sY = sAK. \quad \textbf{(1.84)}$$

where s is the savings rate. It can be concluded from (1.84) that the long term growth rates of capital and income are both equal to:

$$\frac{\dot{K}}{K} = sA. \quad \textbf{(1.85)}$$

This long term result is obtainable even in the short term with the simpler linear production function (the so called "AK" function):

$$Y_t = AK_t. \quad \textbf{(1.86)}$$

That was the approach made by King and Rebelo (1990). The savings rate is endogenously determined by means of a representative consumer maximising a utility function similar to (1.27). However, the fixed savings rate assumption is enough to show that it is the different

[10] Surveys of endogenous growth literature include Barro (1995), Boltho and Holtham (1992), Hammond and Rodriguez-Clare (1993), Romer (1991) and Verspagen (1992).

technology that is responsible for the endogenous growth result.

If a representative consumer maximises an intertemporal utility function as the one expressed in (1.29) and considering that here it happens that r=A, it can be shown that[11]:

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = \frac{\dot{K}}{K} = \theta^{-1}(A - \rho). \qquad \textbf{(1.87)}$$

Two features of the AK model are worth noting:

- There are no transitional dynamics. Adjustment to a change in, say, the savings rate, is immediate. If savings increase as a proportion of income, the growth rate adjusts immediately and permanently to its higher level.

- This model can be seen as the limit of the neoclassical simpler model, when α approaches unity.

None of these models imply convergence of any kind. If two countries have different technologies (different As) or different saving rates (because of different preferences or different tax policies), they will not converge. On the contrary, they will display permanent differences in their growth rates.

b) Growth as a side effect of other activities

A number of authors have developed models that can be classified under this heading. To make the algebra simpler, consider a Cobb-Douglas production function of the form:

$$Y_t = \Omega K_t^{\alpha}(A_t L_t)^{1-\alpha}, \qquad \textbf{(1.88)}$$

where Ω is a constant. Labour is supposed to grow at the rate n. Consider that capital does not depreciate and is accumulated from foregone consumption according to a function G:

$$\dot{K} = G(Y-C,K). \qquad \textbf{(1.89)}$$

[11]See Barro and Sala-i-Martin (1995) for a complete derivation of this result.

Growth will be seen to result either from the accumulation of capital, from an endogenously determined learning process or from the enhancing productivity qualities of public goods. These different engines of growth will translate themselves into different formulations for A_t.

Growth as a side effect of the accumulation of capital

Technical progress is seen as a consequence of the accumulation of capital, according to the following expression:

$$A_t = K_t^{\eta}. \quad \textbf{(1.90)}$$

The economic justification of this last expression is different across authors, as it will become clear soon.

1st case: $0<\eta<1$.

This is the original formulation of Arrow (1962) and Sheshinski (1967). According to the latter author (p. 33), "A_t is assumed to reflect accumulated experience in the production of investment goods." The function G is taken to be simply as G=Y-C. Substitute expression (1.90) into (1.88) take logs and differentiate, to get:

$$\dot{\log Y} = [\alpha + \eta(1-\alpha)]\dot{\log K} + [1-\alpha]\dot{\log L}. \quad \textbf{(1.91)}$$

Note that the production function displays increasing returns to scale, but decreasing returns to any of the factors. There is a unique growth rate that assures that capital and income grow at the same rate:

$$\dot{\log Y} = \dot{\log K} \Rightarrow \dot{\log Y} = \frac{n}{1-\eta}. \quad \textbf{(1.92)}$$

This model has the inconvenient property that makes growth depend on the rate of growth of population. If population growth is zero, we are back to the neoclassical simpler model. Nevertheless, it exhibits an interesting departure from the neoclassical framework: returns to scale are not constant, so that payments to factors according to marginal products do not sum up to total income. Consequently, private decisions drive the economy to a sub-optimal equilibrium, as shown by Sheshinski (1967). Since the next case is a more radical departure from neoclassical assumptions, this one will not be further discussed.

2nd case: $\eta = 1$.

This case is developed by Barro and Sala-i-Martin (1992c) and is inspired in Romer (1986). As before, G=Y-C. Population is supposed to be constant, or, more precisely, this is a per capita model. The interpretation of A_t differs from Arrow' s. Here, it is supposed that:

$$A_t = K_{a,t}, \tag{1.93}$$

where $K_{a,t}$ is the average level of capital used by other producers. Substituting (1.93) into (1.88) the production function becomes:

$$Y = \Omega K^{\alpha} K_a^{1-\alpha}. \tag{1.94}$$

In its original formulation by Romer (1986) K represents accumulated knowledge, or a composite capital good that includes knowledge and physical capital. Each firm observes decreasing returns to *its own* level of capital but returns remain constant to *the total* level of capital. This formalization allows for the existence of a competitive equilibrium that is not Pareto optimal. The private return to capital is equal to:

$$r = \alpha\Omega\left(\frac{K_a}{K}\right)^{1-\alpha} = \alpha\Omega, \tag{1.95}$$

where it was considered that $K=K_a$. It is not difficult to notice that we are back to the AK model, with $A=\Omega$. Accordingly:

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = \frac{\dot{K}}{K} = \theta^{-1}(\alpha\Omega - \rho). \tag{1.96}$$

If the social return to capital (equal to Ω) was considered, the optimal growth rate would be higher and equal to:

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = \frac{\dot{K}}{K} = \theta^{-1}(\Omega - \rho). \qquad \textbf{(1.97)}$$

When private agents make their decisions, they do not take the external benefits of their investment programmes into account. This is the reason why the competitive equilibrium growth rate falls short of the social optimum one.

3rdcase: $\eta > 1$.

This is the case that corresponds to the original Romer (1986) formulation. Aggregate returns to capital are supposed to be higher than 1. The function G(I, K) is assumed to be concave and homogeneous of degree one, so that:

$$\dot{K} = G(Y-C,K) = Kg\left(\frac{Y-C}{K}\right). \qquad \textbf{(1.98)}$$

Romer imposes two other restrictions on g. The derivative function of g obeys the restriction g'(0)=1, and, more important, returns in research are strongly decreasing so that g is limited from above:

$$g\left(\frac{Y-C}{K}\right) < \alpha, \qquad \textit{for all } \frac{Y-C}{K}. \qquad \textbf{(1.99)}$$

This condition assures that the rate of growth of the state variable K remains bounded and therefore allows for the existence of an optimum.

In general, this formulation can produce ever increasing growth rates. Furthermore, growth rates may well depend on the size of the country, so that bigger countries grow faster. This model also displays the already discussed edge between social optimal and competitive equilibrium growth rates.

<u>Growth as a result of learning by studying</u>

In his 1988 paper, Lucas presents two endogenous growth models where growth results from a labour augmenting technical progress that derives either from endogenous decisions on studying or from learning by doing.

Suppose that individuals spend a proportion u of their time working, (1-u) being time spent in studying. Their productivity increases in proportion to time spent studying:

$$\dot{A}_t = A_t \phi (1 - u_t), \qquad \textbf{(1.100)}$$

There ϕ is a constant. Note that u_t is a choice variable. Under the maximisation of the usual intertemporal utility function, the balanced growth path will result in a growth rate that is equal to[12]:

$$\frac{\dot{Y}}{Y} = \frac{\dot{C}}{C} = \frac{\dot{K}}{K} = \theta^{-1}(\phi - \rho). \qquad \textbf{(1.101)}$$

This result is similar to the AK model growth rate. Nevertheless, this model is more complex due to the existence of two state variables (A and K), so that transitional dynamics are at work.

Lucas (1988) first model is a bit more complex. It involves an externality that arises from the learning process. The complete model production function is:

$$Y_t = \Omega K_t^{\alpha} (A_t L_t)^{1-\alpha} A_t^{\gamma}, \qquad \textbf{(1.102)}$$

with A^{γ}_t reflecting this hypothesis. As expected, this externality leads to the conclusion that the market economy will display a lower growth rate, since people will study "less than they should," not considering that their own skill is the other's catalyst.

<u>Growth as a result of learning by doing</u>

The reader may be that rather practical man that thinks people learn more on the job than at school, and that studying is more properly assigned to leisure and has no productive impact. In that case, the previous model is easily modified to fit this alternative specification.

Let:

---

[12]See Lucas (1988) for a complete derivation.

$$\dot{A}_t = A_t \phi u_t, \tag{1.103}$$

so that labour productivity increases with time spent on the job and not on leisure activities. The algebraic results would be much the same.

Lucas (1988) follows this idea in a second model that has the interesting variation of being a two goods and more than one country framework. Its main contents are sketched below.

There are two goods in the world economy: a high technology and a low technology one. Production functions are equal across countries and supposed to be Ricardian (labour is the only factor.) The high technology good production displays a higher learning by doing potential (meaning a higher "ϕ" in terms of equation (1.103)).

If the economies that compose the world economy are open ones, countries will completely specialise in the production of one of the goods. If the goods are good substitutes, Lucas show that countries may find themselves stuck in the production of either high or low technology goods, according to the initial conditions. Due to the lower learning by doing potential of low technology production, low technology countries will grow less than high technology ones.

Growth as a result of the provision of public goods

Barro and Sala-i-Martin (1992c) include three models that can explain growth as a result of the public provision of public goods or services. These are productive inputs to private producers. One model considers public services that are rival but excludable. Another allows for congestion of the public services. The one sketched here treats public services as non-rival and non-excludable. Formally, they are much alike.

Regard labour as constant at the unit value and let:

$$A_t = G_t, \tag{1.104}$$

where G_t is the total amount of public services. These services are financed by an income tax

τ:

$$G_t = \tau Y_t. \tag{1.105}$$

Private agents take the level of public services as given, so that the interest rate is equal to:

$$r = (1-\tau)\alpha\Omega\left(\frac{K}{G}\right)^{1-\alpha}. \tag{1.106}$$

Some algebra permits to derive the following expression for r, from expressions (1.105) and (1.106):

$$r = (1-\tau)\alpha\Omega^{\frac{1}{\alpha}}\tau^{\frac{1-\alpha}{\alpha}}. \tag{1.107}$$

This comes to be another rationale for the "AK" model, with A=r defined as above. If the consumer is maximising the usual utility function (as in Barro and Sala-i-Martin (1992c)) income, capital, public services and taxes all grow at the same rate:

$$\frac{\dot{Y}}{Y} = \frac{\dot{G}}{G} = \frac{\dot{T}}{T} = \frac{\dot{K}}{K} = \theta^{-1}(r-\rho). \tag{1.108}$$

The reader is probably looking for the externality included in this model. It derives from the fact that private agents underestimate the marginal productivity of capital, not considering that an increase in production means an increase in taxes and in the provision of public services. The higher level of public services acts as an effective counterweight to a decreasing productivity of capital. As usual, the "social planner" growth rate is higher.

c) Growth as a result of profit induced R&D

The first group of endogenous growth models previously presented were either straightforward modifications of the neoclassical model, as is the case of the "AK model," or models that display growth as a kind of by-product. All of them deliver important if partial insights.

In the former case, it can be argued that endogenous growth is technically delivered without a convincing economic structure or explanation behind it. Decreasing returns to each factor combined to constant returns to scale are compatible with a perfectly competitive economy. No wonder then that the following question is: is there an alternative economic structure that produces results that fit with an aggregate function equal or close to the "AK" one? One attempt to solve this was made by Romer (1986). He showed that a competitive equilibrium may well coexist with aggregate non decreasing returns if some externalities are at work. Some years later, the same author (Romer (1990a)) and other writers (Helpman (1992), Barro and Sala-i-Martin (1992c)) developed models that settle the problem by dropping the perfect competition hypothesis and adopting a monopolistic competition layout.

All the aforementioned contributions make growth endogenous by explicitly introducing a research and development sector that is profit motivated. Some of these authors[13] refer to a come back to Schumpeterian ideas of creative destruction. This is an important difference from all models considered until now, from a formal but also explicative point of view.

In these models the R&D sector either produces designs of new goods or factors or improves the quality of existing ones. The new or improved factors or goods result in economic growth. These two cases are formally quite close but will be treated separately.

Increasing product or factor variety

Barro and Sala-i-Martin (1992c) model may be seen as a restriction of Romer (1990a). Helpman (1992) and Romer (1990a) are variants of the same idea. This allows for a common exposition that will highlight the main ideas and differences of these approaches.

The production function for final goods is:

$$Y_t = H_{y,t}^{\alpha} L_t^{1-\alpha-\beta} \int_o^{N_t} x_{i,t}^{\beta}\, di, \tag{1.109}$$

where $H_{y,t}$ is the amount of human capital used in the production of final goods, L_t is the amount of labour, and $x_{i,t}$ is the amount of producer durable i. It is assumed that there is a continuum of durables, its number being N_t. Barro and Sala-i-Martin consider a restricted

[13]For instance, Helpman (1992).

version of (1.109), with H_y=1 and α=0. Helpman[14] considers the case where H_y=L=1.

Durables are produced from forgone consumption[15], so that:

$$K_t = \eta\int_o^{N_t} x_{i,t}\, di. \tag{1.110}$$

The usual resource restriction applies:

$$\dot{K}_t = Y_t - C_t. \tag{1.111}$$

Labour and human capital are constant over time. The last may be used in the production of final goods or in the R&D sector:

$$H_{Y,t} + H_{D,t} = H. \tag{1.112}$$

Suppose that each durable is used in the same quantity, irrespective of time:

$$x_{i,t} = \bar{x}. \tag{1.113}$$

If (1.113) holds, the production of final goods is equal to:

$$Y_t = H_{y,t}^{\alpha} L^{1-\alpha-\beta} N_t \bar{x}^{\beta}. \tag{1.114}$$

Using (1.110) and (1.113), it is possible to substitute for N_t in (1.114) and get:

$$Y_t = \eta H_{y,t}^{\alpha} L^{1-\alpha-\beta}\, \bar{x}^{\beta-1} K_t. \tag{1.115}$$

Here is the "AK" model again. If K grows at a rate g, Y will grow at the same rate, if H_y and L stay constant. In other words, there is a balanced growth possibility for this economy. The market structure that supports this path in the steady-state is described next.

[14]Helpman (1992) considers Y to be an index defined in terms of varieties of high-tech products, but this index can easily be reinterpreted as a final good produced by means of varieties of high-tech inputs.

[15]Helpman's version is somewhat different, durables being produced according to a Ricardian production function.

The constant returns to capital result means that there is no perfect competition solution for this model[16]. Instead, a monopolistic competition framework will be adopted.

Consider an economy divided into three sectors: final goods production, production of durables and research and development.

The final goods sector produces according to the production function (1.109). Note that this function implies that the elasticity of substitution between each pair of durables is constant and equal to $(1-\beta)^{-1}$. Demand for durables functions have constant elasticity so that marginal revenue $MR(i)=\beta p(i)$.

There is only one firm that produces and sells durable i. Monopolistic pricing means that marginal revenue is set equal to marginal cost. The marginal cost of producing durables is equal to $r\eta$, r being the interest rate. A symmetry argument make it possible to conclude that prices for each durable are equal:

$$MR_{i,t} = MC_{i,t} \Leftrightarrow \beta p_{i,t} = r_t \eta \Rightarrow \bar{p}_t = \frac{r_t \eta}{\beta}. \tag{1.116}$$

If r is constant in time, the price p and quantities x will also be invariant in time. The operational profits are the same across firms and through time and equal to:

$$\pi = (\bar{p} - r\eta)\bar{x} = (1-\beta)\bar{p}\bar{x}. \tag{1.117}$$

Each firm buys the design for its durable from the R&D sector. In this sector, designs for durables are produced by means of human capital only, N_t being the number of designs produced till time t. It is assumed that:

$$\dot{N}_t = \delta H_{N,t} N_t \tag{1.118}$$

The marginal productivity of human capital equals δN_t and therefore increases linearly with N_t. The more designs have already been invented until today, the more are going to be

[16]At least one that will not involve some kind of externalities.

invented in this period. One wants to formalise the idea that knowledge used in producing old designs becomes public and is useful for producing new designs.

If $w_{H,t}$ is the human capital wage, the price of a design is:

$$P_D = \frac{w_{H,t}}{\delta N_t}. \qquad \textbf{(1.119)}$$

Barro and Sala-i-Martin do not consider equation (1.118) and the implied externality. They make P_D exogenous and constant.

Market for designs is competitive and the price for designs is such that durables producers' profits are driven to zero. If r is constant, this implies that:

$$P_D = \frac{\pi}{r}. \qquad \textbf{(1.120)}$$

Romer (1990a) does not attempt to study the behaviour of this model off the steady-state. Instead, he provides and compares two steady-state solutions: one where there is monopolistic competition in the supply of durables as described above, and other that corresponds to the social optimum. The utility function is the usual one with constant marginal utility elasticity.

In the first situation, income, the number of designs and capital are shown to grow at the same rate g_1:

$$g_1 = \frac{\delta H - \Lambda\rho}{\rho\Lambda + 1}, \qquad \textbf{(1.121)}$$

where:

$$\Lambda = \frac{\alpha}{(1-\alpha-\beta)(\alpha+\beta)}. \qquad \textbf{(1.122)}$$

Romer also shows that social optimum balanced growth rate is equal to g_2 with:

$$g_2 = \frac{\delta H - \Theta\rho}{\rho\Theta + 1(1-\Theta)}. \qquad (1.123)$$

where:

$$\Theta = \frac{\alpha}{\alpha+\beta}. \qquad (1.124)$$

It results that $g_2 > g_1$, meaning that the growth rate under monopolistic competition is smaller than the optimal one.

In fact, two market failures make that too little human capital is devoted to research. Research has some effects that are external to the engaged individuals, and it produces an input that is purchased by a sector that practises monopoly pricing. An individual that produces research raises the productivity of all the other researchers. This externality is not accounted for in the price for designs. Also, the producer of a design receives only a fraction of the corresponding marginal output, due to the monopoly structure of the durables production sector.

It is clear from the above discussion that this model does not imply any kind of convergence between economies. Even when preferences are the same (ρ is the same in every economy), growth will generally be different, depending on the stock of human capital H, and on adopted policies. These may get the growth rate closer to its social optimum value (the policies would involve subsidising the accumulation of knowledge.) In general terms, one would expect that:

- countries that have a higher stock of human capital will grow faster;
- if two or more countries completely integrate their economies, they will grow faster, since they will share a higher amount of human capital, equal to the sum of their previous ones. This model predicts that integration into an economy with a large stock of human capital will boost growth.

Increasing product or factor quality

The models presented above are subject to a natural criticism: once invented, products/factors are used forever. One could think that a more realistic approach would be to allow for *quality* improvements in goods or factors, instead of *quantity* improvements. In practice, one would expect both phenomena to occur. When colour TV was invented people substituted colour TV sets for black and white ones. Nevertheless, black and white sets are still made and sold at a lower price. Helpman (1992) presents a quality improvement model formalised in lines not very different from the quantity increasing model presented above, the R&D sector being responsible for the design of product improvements and not of new products.

1.5 Convergence and divergence in growth theory: a synthesis

Three different approaches to growth were reviewed: the neoclassical exogenous growth approach, the technological catch up approach and the endogenous growth paradigm. As stated, they seem to imply different results for the convergence of incomes per head:

- The neoclassical models usually lead to a conditional convergence result.

- The technological catch up model allows for the existence of a limited convergence outcome. Only countries with the necessary social capability will eventually converge to the income path of a leading country.

- Endogenous growth models suggest a tendency for divergence of incomes. Growth rates usually depend on country specific parameters or policies.

In fact, this is not a clear-cut image of the literature on the subject. If this was so, empiric methods could well discriminate between these three "families" of models, provided sufficiently powerful methods were available. For example, if the data showed a tendency for convergence, endogenous growth models would be put in doubt[17].

It is possible to justify the existence of several "convergence clubs" within the neoclassical framework. One only needs to drop the assumption of a universal exogenous growth rate and postulate that there are different groups of countries that converge to different steady-states[18].

---

[17]This seems to be the reasoning of some authors, such as Mankiw, Romer and Weil (1992).

[18]This is the idea of Durlauf and Johnson (1992). In chapter 2, their empirical work derived from this approach is surveyed.

Moreover, it is possible to reconcile endogenous growth ideas with convergence and technological catch up. Easterly, Kremer, Pritchett and Summers (1993) and Lucas (1993) provide the clue. Suppose that a leading country generates the rate of technological progress according to one of the endogenous models presented earlier in this chapter. Progress in the leading country spills over to followers. The pace of the adoption of new technologies may depend on country specific characteristics and policies. This results in a diffusion model where growth is endogenous for the leading country or countries but exogenous for the followers. Policies in following countries would only determine their transitional dynamics and their relative position to the leader in the steady-state[19].

2. Convergence of international factor prices

In two important papers, Samuelson (1948, 1949) argued for the complete equalisation of international factor prices under a restrictive set of assumptions, in what became later known as the FPE (factor prices equalisation) theorem. The main assumptions for the theorem to hold are the following[20]:

i) There is complete arbitrage in the international goods market and no tariffs or transport costs, so that prices are the same in each country;

ii) Technologies (production functions) are the same in each country and factors are qualitatively the same;

iii) Production functions exhibit constant returns to scale and diminishing marginal productivities;

iv) The number of factors is less than or equal to the number of goods;

v) No factors of production can move between countries;

---

[19]Other diffusion models are presented by Parente and Prescott (1991) and Jovanovic and Lach (1991).

[20]See Samuelson (1948,1949) and Burgman and Geppert (1993).

vi) Tastes are the same;

vii) Countries are sufficiently close together in factor proportions.

The following "factor-price equalisation model" is borrowed from Dollar, Wolff and Baumol (1988) and illustrates Samuelson's insights.

Let each country produce n goods with n factors. Production functions are continuous and homogeneous of degree one. Each element a_{ij} of a matrix A represents the requirements of factor i to produce good j. Each element of A depends on relative factor prices. If w and p are respectively the vector of factor and goods prices, zero profits imply that:

$$Aw = p. \qquad \textbf{(1.125)}$$

Suppose that A is invertible for all w. Then,

$$w = A^{-1}p. \qquad \textbf{(1.126)}$$

If A is equal for every country (and this is verified if every country has the same technology) and if goods prices p are the same everywhere, then factor prices will be equal.

In practice, one would not expect the restrictive assumptions of the FPE theorem to be verified. Indeed, if they were all verified, FPE would certainly occur, since this is a mathematical demonstrated theorem. One can nevertheless expect that some approximations to the theorems' hypothesis actually take place so that some tendency for equalisation arises. For instance, it is sensible to expect some sort of convergence between factor prices in not very different economies that are engaged in a process of economic integration like the European Union[21].

The fact that equality between factor prices is not expected means that the insinuated pattern of convergence is not point wise. There is even scope for conditional convergence, if different country conditions (as taxation) imply permanent differences between prices for the same factor in different countries.

[21]This is the theme of two papers (Tovias (1982) and Gremmen (1985)) to be commented later in chapter 3, where actual processes and results on measuring convergence are dealt with.

3. European Economic and Monetary Union and nominal convergence

From the late eighties on, the twelve countries that formed the European Community engaged themselves in a process of further integration that may well drive (some of) them into a full Economic and Monetary Union by the end of this century. Two important events marked this evolution: the signature of the Single European Act in 1986 and the agreement on a Treaty on European Union in 1991[22]. The latter is commonly referred as the Maastricht Treaty, from the Dutch town where it was agreed.

The main purpose of the European Single Act was to complete the unification of national markets by the end of 1992. From then on, complete freedom of movements for goods, services, persons and capital were to be assured.

The Maastricht Treaty was and still is a much discussed step forward. It includes some provisions on political union and establishes the stages for the achievement of a monetary union that includes the establishment of a European Central Bank (ECB) and of a single currency. This treaty involves several criteria of economic convergence in the constitution of the European Monetary Union (EMU, for short.)

The transition into a full EMU is to be accomplished in three stages. The first stage consisted mainly in the adoption of multi-annual programmes to ensure enduring convergence indispensable for EMU, specifically in what concerns fiscal discipline and price stability together with a complete liberalization of capital movements.

Stage two began in January 1994, a European Monetary Institute (EMI) being created. Its main task is to help national central banks in the coordination of monetary policy and to plan the move to the third phase (monetary union.) Some provisions are made for this stage not to last indefinitely, according to article 109:

- The European Council should, by the end of 1996, decide by qualified majority if a majority

[22]Belmont European Policy Centre (1991) includes the text of the Treaty. Gros and Thygesen (1992) and De Grauwe (1994) cover the economics of EMU.

of countries fulfil the "necessary conditions" for the adoption of a single currency, if it is appropriate to begin the third stage, and if so set a date for it.

- If by the end of 1997 no date for the beginning of the third stage has been agreed upon, than it will start in the beginning of 1999, with whichever members are ready, even if they form a minority of countries.

The "necessary conditions" referred to above are detailed in two protocols annexed to the Maastricht treaty: the protocol on "the convergence criteria" and the protocol "on the excessive-deficit procedure."

There are four convergence criteria to which a country should abide in order to participate in the EMU:

1) The *price stability criterion* establishes that the average rate of inflation during one year before examination should not exceed by more than 1.5 percentage points that of at most the three best performing countries in terms of price stability[23].

2) The *government budgetary positions criterion* stipulates that the ratio of government deficit to GDP should not exceed a reference value (3 per cent, in the annexed protocol) and that public debt should be no more than 60 per cent of GDP (again a value defined in the protocol.) If reference values are exceeded, the ratios should be diminishing at a satisfactory pace. Exceptional and temporary deficits above 3 per cent are not considered a violation of this criterion.

3) The *ERM participation requirement* specifies that a country should be an ERM participant during at least two years with its currency not being a focus of a major tension, namely not being subject to any devaluation.

4) An *interest rate criterion* determines that a currency long-term interest rate on government bonds should not exceed by more than two percentage points the similar interest rate in at

[23]In this context, "price stability" means a low inflation rate.

most the three best performing member states in terms of price stability.

If one compares each of these criteria with the general definitions of convergence one may conclude that:

i) If a country inflation rate converges unconditionally to the inflation rate of the third best performer and if this convergence, even if not point-wise, displays a sufficiently small variance, then that country should obey the price stability criterion from a certain point in time on. If the referred convergence is conditional, the implied constant should be smaller than 1.5 percentage points.

ii) Similar statements to i) apply to the interest rate criterion[24].

iii) Budget positions criteria are a bit more difficult to match with convergence definitions. This happens because of the one sided nature of them. Once deficits or public debts do not exceed certain values, the criteria are verified and there is no need for any kind of convergence between series. However, if a country's deficit or debt are above the reference values but converging to a stationary series with a mean lower than those values, there is a point in time such that from then on the criteria are fulfilled.

iv) The ERM participation requirement may be regarded as a requirement for unconditional convergence of exchange rates. If exchange rates have converged so that the underlying variance is sufficiently small, then countries will have no problems in staying inside the ERM without causing major tensions[25].

---

[24]Irrespective of the Maastricht Treaty, there are reasons to expect convergence of interest and inflation rates (at least for tradable goods) in a fixed exchange rate regime, due to arbitrage in financial markets and to the law of one price, respectively (See Honohan (1992), Karfakis and Moschos (1990) and Pigott (1994)). However, the European Union is not (yet) in a complete fixed exchange rate regime. The "Maastricht convergence criteria" are pre-conditions for the feasibility of a monetary union, and not a consequence of it.

[25]The recent enlargement of ERM bands means that the aforementioned variance may be quite large...

Chapter 2

Measuring Convergence: Methods and Results

Introduction

Chapter 1 contained a review of literature on the convergence issue in Economics. The different models surveyed there and their sometimes conflicting results gave way to an ever growing literature on the empirics of convergence. Different tests and procedures have been suggested by different authors, and sometimes differing outcomes arise from the application of different methods to the same or similar data sets. This chapter is both a tentative to identify the main methods used to measure convergence and to summarise the more important results[1].

The first section in this chapter surveys results using one of the more popular techniques of measuring convergence: the analysis of the time series of a cross-section dispersion measure. This technique has been applied to different fields of Economics.

As shown in chapter 1, convergence issues are present in different parts of economic theory and practice. Nevertheless, recent theoretical and empirical developments led to a surge in the literature devoted to convergence in incomes per head across different economies. Methods that deal with incomes per head and its results are reviewed in section 2.

A test for convergence using the Kalman filter is presented in section 3. This method has been applied to nominal variables in the European Community but can be used in other settings.

In section 4 some recent literature on the identification of supply and demand shocks using vector autoregressive models is examined and its relationship to the convergence definitions presented in chapter 1 assessed.

Finally, section 5 deals with a somehow different but related kind of convergence: the

---

[1]Pack (1994) surveys some of the empirical evidence on endogenous growth with some references to convergence. Sala-i-Martin (1994) reviews the convergence evidence that results from initial value regressions.

convergence of a vector of several variables.

1. Dispersion measures methods

These methods are based in a simple idea: if n time series are converging, their dispersion should be declining over time. They usually imply the computation of a time series of one or several dispersion measures and a more or less rigorous analysis of their declining tendency. Applications of this idea may be found in different fields of the applied economics literature[2].

It was shown in the introduction to chapter 2 that convergence of series according to the definition there included is compatible with both a declining and an increasing variance for the cross-sectional distribution of the series. Nevertheless, a declining cross-section dispersion is usually a sign of unconditional convergence, even according to the aforementioned definitions if the series start sufficiently apart. This is one possible interpretation for the frequent use of this kind of procedure[3].

1.1 Dispersion measures applied to growth theory

Baumol and Wolf (1988) studied convergence of GNPs per capita in a sample of 19 European countries. On the whole, their paper supports the idea of "local" or "limited" convergence. They analyse the evolution of the coefficient of variation of GNP per capita from 1870 to 1913 for the top 8, 9, 10, and 11 countries in 1870. From visual inspection of the series, they find some evidence of convergence in what concerns the top eight countries, since the 9th time series is not declining.

The same procedure is applied to real GDPs from 1950 to 1980 in a larger sample of countries. The time path of the coefficient of variation for each year was calculated for the top 10, 12, 14, 16, 18, 22, 24, 26, 28, 30, 35, 40, 45, 50, 55, and 60 countries in what concerns

---

[2]For example, Baumol and Wolf (1988), Lichtenberg (1994) and Barro and Sala-i-Martin (1995) apply this to income per capita and Tovias (1982) and Mohktari and Rassekh (1989) use the same idea when studying the convergence of international factor prices.

[3]The fact that the researchers are working with different concepts of convergence is of course another interpretation. Lichtenberg (1994) explicitly defines convergence as a tendency for the cross-section variance to decrease through time.

real GDP per head. According to the authors, "there is a sharp break in the pattern of behaviour between the samples that include fewer than 16 countries and those that include 16 or more." All countries together, excluding less developed ones, also show some convergence, larger samples not displaying any.

Barro and Sala-i-Martin (1995) computed a time series of the cross-sectional variance of the log of per capita GDP or personal income across regions within Germany, the United Kingdom, Italy, France, and Spain, between 1950 and 1990, across Japanese prefectures, from 1930 to 1990, and across US states, between 1880 and 1990. There is a tendency for this measure to decline in the beginning of the period and to become more or less constant towards the end, sharp increases being usually absent. Note that the same authors reported absolute convergence results for the same series (these results will be discussed later.)

Neven and Gouyette (1995) report similar measures for the output per head of 172 European Community regions for the period 1980-1989. A declining pattern is detected from 1984 on. Dividing the sample into Northern and Southern regions, the authors concluded that there is a tendency for the former to become more alike and for the latter to diverge.

Lichtenberg (1994) applies the F-test for the equality of variances to the convergence framework. Let σ_0^2 and σ_1^2 denote respectively the initial and final period cross-sectional variance, respectively. Under the null hypothesis that they are equal, the ratio R of the sample estimators follows an F(n-2, n-2) distribution, where n is the number of series. In an application to 22 OECD countries in the post-war period, Lichtenberg found that R=1.57, and could not dismiss the no convergence null. That result was compatible with a highly significant negative coefficient when growth rates were regressed on initial values in what constitutes an interesting empirical illustration of the differences between "beta" and "sigma" convergence.

1.2 Testing factor price convergence using dispersion measures

Tovias (1982) tested factor price convergence in the European Community (only founding members were considered) using dispersion methods. The EC was chosen because:

- it was a customs union that supposedly led to the establishment of common prices of goods, but for the existence of transport costs;

- the period covered (1950 to 1973) included a number of years before the EC was created in 1958, a transitional period (January 1958-July 1968) and a post-integration period;

- the economic conditions within and between countries were not totally at odds with the hypothesis of the FPE theorem.

The chosen dispersion measures were the standard deviation and the coefficient of variation. According to the author, if a decrease over time is detected in both, there is evidence of "strong convergence." If both measures increase, there is divergence. If the only declining measure is the coefficient of variation, "relative convergence" is at work.

These tests were performed in three different data sets, the differences between them resulting from different sources, included countries[4] or period considered. The results were not very different, and the author concluded that:

"(...) from the different tests relative to five (or four) original members of the EEC it emerges that convergence in labour costs took place between the mid-50s and the mid-60s, a decade in which at least in two sub-periods (1957-59 and 1961-64) there is some evidence of strong convergence in labour costs. From 1968 on, the trend clearly reversed, although it seems that up to 1970, the dispersion of labour costs in the EEC was still smaller in relative terms than before its creation."[5]

A two-country version of this test consists in considering the time series of differences and ratios of labour costs between countries. Movements in the ratio series in the direction of unity, or in the differences series in the direction of zero, are taken as evidence of convergence. When appliying this to France and Germany between 1950 and 1979, the author finds evidence of convergence between 1956 and 1968 only.

---

[4]Luxembourg was never considered.

[5]*in* Tovias (1982), p. 387.

The coefficient of variation test is applied to an extended set of 16 OECD countries between 1961 and 1984 by Mohktari and Rassekh (1989)[6]. The coefficient is pictured in a graph and its declining pattern is taken as evidence of convergence.

A more elaborate technique consists of identifying explanatory variables that account for the time and/or cross country behaviour of dispersion measures. This is the approach of Gremmen (1985) and Mohktari and Rassekh (1989). All these authors use linear regression methods.

Gremmen considers the following equation:

$$\ln(\frac{w_i}{w_j})_t = constant + \alpha \ln(V_{ij})_t + \beta \ln(\frac{K_i/L_i}{K_j/L_j})_t, \qquad \textbf{(2.1)}$$

where w_i is the labour cost in country i, V_{ij} is a measure of trade involvement between countries i and j, and K_i/L_i is the physical capital labour ratio in country i.

Gremmen considers that the predictions of the FPE theorem imply that α is positive and β close to zero between countries with high trade involvement. If trade involvement is low, α should be close to zero and β positive: when approaching an autarky situation, differences in wages are a reflection of different technologies.

The first test is carried out using a high trade involvement group of countries. These are the original members of the EC except Luxembourg between 1959 and 1979. The second test sample consists in 26 countries scattered around the world, considered a low trade involvement group. Cross-section and time series are pooled, and the results agree with the hypothesis.

Mohktari and Rassekh (1989) propose an equation for the coefficient of variation of labour costs (CVW) in their already mentioned sample of 16 OECD countries. The log of CVW is considered a log-linear function of the coefficient of variation of the capital-labour ratio, a trade openness measure and the employment rate. This last variable is subsequently not found significant, but the other two exert a positive influence on labour costs convergence.

[6]These countries are Canada, the USA, Denmark, Holland, Sweden, Japan, New Zealand, Austria, Belgium, Finland, Ireland, France, Norway, Switzerland and the United Kingdom.

2. The empirics of economic growth and convergence

Some of the methods for measuring convergence were devised for coping with the "convergence debate" in the empirics of economic growth and usually are not found in other applications. These methods are presented together under this heading.

2.1 "Initial value" methods

Consider a cross-section of economic non-stationary series. If they are converging, then the ones that started from a level that is further from the steady-state should be growing faster. Here, "initial value methods" are methods that exploit this idea.

a) Unconditional convergence and simple linear regression

The inverse relationship between growth rates and initial values was estimated in linear form by Baumol (1986) in what became a much cited result:

$$g = 5.25 - 0.77 \ln(GDP/WH_{1870}), \quad R^2 = 0.88. \tag{2.2}$$

This regression was performed using data on 16 industrialised countries, g is the average growth rate from 1870 to 1979, and WH_{1870} being the number of work hours in 1870. The regression above does not produce good results when applied to a larger set of countries. Convergence does not appear ubiquitous, and Baumol writes about the existence of convergence clubs. This point will be developed later.

The rationale for this and similar regressions may result either from a neoclassical exogenous growth reasoning or from a technological catch up process.

The neoclassical exogenous growth empirical framework is provided by Barro and Sala-i-Martin (1992a, 1992b, 1995) and Mankiw, Romer and Weil (1992). Recall equation (1.60). Consider a time length equal to one and include a random disturbance (i is a country index):

$$\frac{\log Y_{it}}{L_{it}} - \log\frac{Y_{i,t-1}}{L_{i,t-1}} = g - (1-e^{\lambda_1})\log\frac{Y_{i,t-1}}{L_{i,t-1}} + (1-e^{\lambda_1})(\log y_i^* + A_{i0}) + u_{it}. \qquad \textbf{(2.3)}$$

Table 2.1
Unconditional convergence results

Sample and Authors	Variable	Estimated λ_1 (estimated st. deviation)	R^2	Estimated σ
98 countries, 1960-85 (B & S, 1992a)	GDP per head	0.0037 (0.0018)	0.04	0.0183
20 OECD countries, 1960-85 (B & S, 1992a)	GDP per head	-0.0095 (0.0028)	0.45	0.0051
47 US states 1880-1990 (B & S, 1995)	personal income	-0.0174 (0.0026)	0.89	0.0015
98 non-oil countries, 1960-85 (M, R & W, 1992)	GDP per working person	0.00360 (0.00219)	0.03 (adjusted)	0.44
75 intermediate countries, 1960-85 (M, R & W, 1992)	GDP per working person	-0.00017 (0.00218)	-0.01 (adjusted)	0.41
22 OECD countries, 1960-85 (M, R & W, 1992)	GDP per working person	-0.0167 (0.0023)	0.46 (adjusted)	0.18
47 Japanese prefectures, 1930-90 (B & S, 1995)	personal income	-0.0279 (0.0033)	0.92 (0.0019)	0.0019
141 EC regions, 1980-89 (N & G, 1995)	GDP per head	-0.053 (0.021)	0.04	n.a.

Notes:
- "B & S" , "M, R & W" and "N & G" stand for "Barro and Sala-i-Martin" , "Mankiw, Romer and Weil," and "Neven and Gouyette", respectively.
- The dependent variable is not the same: Mankiw, Romer and Weil take the differences between logs of incomes in the end and beginning of the period (cumulated growth) while the other authors take the average growth rate.

This last equation implies the following one, for the average growth rate between time 0 and time T:

$$\frac{\log\frac{Y_{iT}}{L_{iT}} - \log\frac{Y_{i0}}{L_{i0}}}{T} = g + \frac{1-e^{\lambda_1 T}}{T}(\log y_i^* + \log A_{i0}) - \frac{1-e^{\lambda_1 T}}{T}\log\frac{Y_{i0}}{L_{i0}} + v_{iT}. \qquad \textbf{(2.4)}$$

v_{iT} is a random disturbance resulting from an weighted average of the u_{it}s. If the steady-state and initial conditions expressed in the A_0s are the same across economies, then equation (2.4) simplifies to the following one:

$$\frac{\log\frac{Y_{iT}}{L_{iT}} - \log\frac{Y_{i0}}{L_{i0}}}{T} = constant - \frac{1-e^{\lambda_1 T}}{T}\log\frac{Y_{i0}}{L_{i0}} + v_{iT}. \quad \textbf{(2.5)}$$

Baumol's work (1986) may be regarded as an *avant la lettre* estimation of (2.5), but Barro and Sala-i-Martin (1992a, 1992b, 1995) provide their own. They estimated equation (2.5) for three groups of countries or regions: a group of 98 countries, the group of the 20 OECD founding countries and a set of 48 US states. Some of their results are reproduced in Table 2.1, along with similar estimations by Mankiw, Romer and Weil (1992) and by Neven and Gouyette (1995).

These results are broadly comparable to Baumol's. There is no evidence of (unconditional) convergence for large samples, but results are more encouraging when more similar economies are considered. As it will be seen shortly, these results may be improved if other variables are introduced to account for conditional convergence.

Comparing growth rates in different groups of countries

Table 2.2
Comparing growth rates in different groups of countries

	1950-60	**1960-73**	**1973-85**	**1950-85**
Richer countries: average growth (std. dev.)	2.3 (0.8)	3.1 (0.6)	1.5 (0.6)	2.3 (0.5)
Poorer countries: average growth (std. dev.)	4.2 (1.5)	5.1 (1.6)	1.9 (0.8)	3.7 (0.9)
T-statistic	3.7	4.1	1.3	4.6

If unconditional convergence is taking place, then poorer countries should be growing faster than richer ones. This hypothesis is directly tested by Dowrick and Nguyen (1989), considering the 24 OECD countries. These authors split the sample in two halves, after ranking countries by income per capita in the beginning of the period. After computing the respective average growth rates, a t-statistic is calculated to test for equality of means. Their results are reproduced in Table 2.2.

The 5 percent critical value for the t-statistic is 2.13, and one can conclude convergence was working in all periods considered except 1973-85, the authors noting that these results were robust to the exclusion of Japan.

b) Conditional convergence and multiple linear regression

The results mentioned until this point suggest that there is some evidence of unconditional convergence in the OECD countries and in the US states, but that there is no sign of it when a more complete sample is considered. Theory nevertheless suggests that conditional convergence could still happen.

The theoretical basis for conditional convergence comes from the exogenous growth neoclassical model, if one allows for steady-state differences in equation (1.60). Similarly, the extended model with human capital provides a rationale for conditioning. In equation (1.73), income per head growth depends on initial income when other variables are also included.

The technological catch up model of Dowrick and Nguyen (1989) also provides a conditional convergence result, as expressed in equation (1.79). The time average growth rate depends on initial income, conditioned on the growth rates of capital and labour.

In different works, Barro (1991) and Barro and Sala-i-Martin (1992a, 1992b, 1995) present conditional convergence regressions for the US states, for 98 non-oil countries, for 20 OECD economies and for several regions of Europe (from Germany, the United Kingdom, Italy, France, the Netherlands, Belgium, Denmark and Spain.) They estimate equations like the following one:

$$\frac{\log\frac{Y_{iT}}{L_{iT}}-\log\frac{Y_{i0}}{L_{i0}}}{T} = constant + linear\ combination\ of\ conditioning\ variables - \frac{1-e^{\lambda_1 T}}{T}\log\frac{Y_{i0}}{L_{i0}} + v_{iT} \quad (2.6)$$

Neven and Gouyette (1995) also estimated this equation for 141 European regions.

Mankiw, Romer and Weil (1992) and Knight, Loyasa and Villanueva (1993) provide comparable research, their equations being:

$$\log\frac{Y_{iT}}{L_{iT}}-\log\frac{Y_{i0}}{L_{i0}} = constant + linear\ combination\ of\ conditioning\ variables - (1-e^{\lambda_1 T})\log\frac{Y_{i0}}{L_{i0}} + v_{iT} \quad (2.7)$$

Conditional convergence may be regarded as multiple regressions where initial income is one of the dependent variables and cumulated or average growth of income is the independent one.

Some of these results are summarised in Table 2.3 .

The conditioning variables differ in these approaches. Mankiw, Romer and Weil derive theirs directly from their extended Solow model. Accordingly to equation (1.73), they consider:

- the average ratio of investment to GDP during the whole period;
- the average percentage of the working age population in secondary school for the period 1980-85, to proxy for human capital investment;
- the average population growth rate during the period considered, $g+\delta$ being considered equal to 0.05 in every country.

Knight, Loyasa and Villanueva consider the same conditioning variables as Mankiw, Romer and Weil, plus the average ratio of general government fixed investment to GDP and the weighted average of tariff rates on imported intermediate and capital goods.

In what concerns Barro and Sala-i-Martin, they take the following:

- primary and school enrolment rates in the beginning of period;

- the average ratio to GDP of government consumption expenditure excluding defence and education from 1970 to 1985;

- the average number of revolutions and coups per year from 1960 to 1985;

Table 2.3

Conditional convergence results

Sample and Authors	**variable**	**Estimated λ_1 (estimated st. deviation)**	**R^2**	**Estimated σ**
1. 98 countries, 1960-85 (B & S, 1992a)	GDP per head	-0.0184 (0.0045)	0.52	0.0133
2. 98 countries, 1960-85 (M, R & W, 1992)	GDP per working person	-0.0137 (0.0019)	0.46 (adjusted)	0.33
3. 81 countries, 1960-85 (K, L & V, 1993)	GDP per working person	-0.0499 (0.006)	n.a.	n.a.
4. 59 developing countries, 1960-85 (K, L & V, 1993)	GDP per working person	-0.2301 (0.0243)	n.a.	n.a.
5. 20 OECD countries, 1960-85 (B & S, 1992a)	GDP per head	-0.0203 (0.0068)	0.69	0.0046
6. 22 OECD countries, 1960-85 (M, R & W, 1992)	GDP per working person	-0.0203 (0.0020)	0.65 (adjusted)	0.15
7. 75 intermediate countries, 1960-85 (M, R & W, 1992)	GDP per working person	-0.0182 (0.0020)	0.43 (adjusted)	0.30
8. 141 EC regions, 1980-89 (N & G, 1995)	GDP per working person	-0.0111 (0.00302)	0.26	n.a.

Notes:
- "B & S," "M, R & W", "K, L & V" and "N &G" stand for "Barro and Sala-i-Martin," "Mankiw, Romer and Weil,", "Knight, Loyasa and Villanueva," and "Neven and Gouyette", respectively.
- The dependent variable is not the same: Mankiw, Romer and Weil and Knight, Loyasa and Villanueva take the differences between logs of incomes in the end and beginning of period (cumulated growth) while the other authors take the average growth rate.
- Knight, Loyasa and Villanueva use a panel data estimating procedure, finding evidence of significant country-specific effects.

- the average number of political assassinations per capita per year from 1960 to 1985;

- the average deviation from unity of the Summers and Heston (1988) purchasing power parity ratio for investment in 1960.

Finally, Neven and Gouyette took country dummy variables as proxies for different country-specific steady-state levels.

If one does not consider the estimates by Knight, Loyasa and Villanueva, the similitude between estimated speeds of convergence is striking. When the two other groups of authors work with more or less the same samples, this result is not so surprising: although the conditioning variables are not the same, they are probably quite correlated. What strikes most this reviewer is the similitude of estimated λs across samples: a value close to 0.02 appears pervasive.

Apparently, allowance for country-specific effects results in a higher estimated speed of convergence, but this could also be the result of including other conditioning variables.

Another conditioning regression is presented by Dowrick and Nguyen (1989). Basing their study on equation (1.79) they provide the following estimated equation, the dependent variable being the relative to US trend growth rate in 23 OECD countries between 1950 and 1985:

$$\hat{q} = constant - \underset{(8.20)}{2.01}\,\hat{y} + \underset{(4.04)}{0.58}\,\hat{l} + \underset{(2.73)}{0.064}\,\hat{k}, \tag{2.8}$$

the dependent variables being the relative initial income per head, the relative rate of growth of employment and the average ratio of gross investment to GDP. Figures in parenthesis are t-statistics. Interestingly, in this paper the author derives the following result for the speed of convergence:

$$\hat{\lambda} = 0.024. \tag{2.9}$$

This results are extended by Dowrick (1992). He splits a sample of 113 countries between 1960 and 1988 into three groups: a "high productivity group," the "middle income economies" and the "poor economies." Conditioning real GDP growth on the growth of the work force and on the average investment rate, he finds evidence of convergence within the richer and the poorer economies, and across economies.

Quah (1993b) criticises initial value regressions on the basis that they are not a good way of measuring convergence because they are subject to the "Galton's fallacy". In an appendix to this chapter, it is argued that this criticism results from the definition of convergence this author adopts.

c) Local versus global convergence

In all initial value methods discussed up to this point, a negative estimated coefficient on initial income is considered as evidence of *global* convergence. Nevertheless, as Durlauf and Johnson (1992) defend, there is a broad scope in theoretic growth models for *local* convergence. Different groups of countries may be characterised by different steady-state growth rates. Here, an initial value regression could still result in a negative initial value coefficient, and the researcher would wrongly conclude in favour of (global) convergence. In other words, the initial value type of test is not very powerful when some sensible alternative hypotheses are considered.

To understand the argument of Durlauf and Johnson, consider that there are M different convergence clubs, the following growth equation being the "true" underlying data generation process:

$$\log\frac{Y_{iT}}{L_{iT}} - \log\frac{Y_{i0}}{L_{i0}} = c_i + \textit{linear combination of conditioning variables} + \beta\log\frac{Y_{i0}}{L_{i0}} + v_{iT}. \qquad \textbf{(2.10)}$$

where c_i takes one of M different values, according to the country's long run equilibrium.[7] It is very likely that the imposition of the same constant across economies results in a negative estimated β that leads to a "global convergence" conclusion.

The authors devise several different ways of identifying convergence clubs and its members, using basically the same data sets and conditioning variables as Mankiw, Romer and Weil (1992). They conclude that (p. 25) "the behaviour of national growth rates in the post war

[7]For the sake of argument, it is considered that the data generation process is similar for every economy, except for the constant.

period is quite compatible with a multiple equilibrium perspective."

Their techniques and results are briefly described below. Two types of equations were estimated: an "unrestricted" one, similar to equation (2.10), and a restricted one, i.e., one that imposes restrictions implied by the human capital augmented Solow growth model.

Method 1: specification tests for different regimes

Countries are split according to initial incomes and literacy rates. Estimated parameters are statistically different. This evidence supports local convergence and is found robust to the addition of other control variables.

Method 2: single control variable estimates

Quandt's (1958) methods are adopted. After exogenously choosing the control variable and the number of splits, their location is set to maximise the likelihood function of the data. According to the authors, "using both income levels and literacy rates to identify different regimes, we have found substantial evidence of heterogeneity in production technologies and of local rather than global convergence in national economies. However, "the characteristics of the regimes differ according to which variable is used to split the sample" (p.20.)

Method 3: regression tree estimates

Using a regression tree method[8], four groups of countries are identified: low income, intermediate income but poor literacy, intermediate income but high literacy, and high income. The authors think their results (p. 23) "are strongly consistent with the view that different economies have access to different aggregate technologies."

In a different but related strand Baumol and Wolf (1988), Chaterji (1992) and Alam and Naseer (1992) also argued in favour of the existence of convergence clubs. This result is usually achieved by adjusting a quadratic function of initial income to the cross-sectional

[8]This method is explained in an appendix to Durlauf and Johnson (1992). Their source is Breinan, Friedman, Olshen and Stone (1984).

growth rates:

$$growth = constant + k_1.(income) + k_2.(income)^2. \qquad \textbf{(2.11)}$$

Baumol and Wolf, using 72 countries from 1950 to 1980, estimated a positive k_1 and a negative k_2. Consequently, growth depends negatively on income only if it is comprised in a specific range. Otherwise, those two variables are negatively correlated. From the specific values for the parameters, they concluded that richer countries were converging to a higher level of income when compared to poorer nations.

2.2 Markov chains and related methods

a) Markov chains and unconditional convergence

The modelling of social and income mobility using Markov chains ideas is not new.[9] Quah (1993a, b and c) develops this methodology and applies it to the cross-section dynamics of economic growth. The main aspects of his methods and results are presented below.

Consider the following law of motion for the distribution of incomes across countries[10]:

$$F_{t+1} = M.F_t, \qquad \textbf{(2.12)}$$

where F_t is a vector that denotes the distribution of incomes, M being a time invariant Markov transition matrix.

In Quah (1993a) F_t is a vector where the ith element is the proportion of countries that fall inside a prespecified quantile. The quantiles resulted from the partition of incomes into five groups defined by a grid of 1/4, 1/2, 1, and 2. The first quantile is the number of countries whose income is smaller than 1/4, the second one is the number of countries whose income is higher than 1/4 but smaller than 1/2, and so forth.

---

[9]See Kemeny and Snell (1960), chapter VII, for an interesting example of its application to social mobility in England and Wales.

[10]"Incomes" should be taken to be "incomes per head as a proportion of the average cross-section income per head."

Considering 118 countries and annual data from 1962 to 1985, Quah estimated two Markov transition matrices: the first one considering a one-year period transition and the other considering a 23-year transition. This last estimate is reproduced below:

$$\hat{M} = \begin{pmatrix} 0.76 & 0.12 & 0.12 & 0.00 & 0.00 \\ 0.52 & 0.31 & 0.10 & 0.07 & 0.00 \\ 0.09 & 0.20 & 0.46 & 0.26 & 0.00 \\ 0.00 & 0.00 & 0.24 & 0.53 & 0.24 \\ 0.00 & 0.00 & 0.00 & 0.05 & 0.95 \end{pmatrix} . \qquad \textbf{(2.13)}$$

In this matrix, element m(i,j) corresponds to the estimated probability for a country in quantile i to swap into quantile j.

An estimate for the long run distribution of cross-country incomes is:

$$\hat{F}_{\infty} = \lim_{s \to \infty} \hat{M}'^{s} \cdot F_{t}. \qquad \textbf{(2.14)}$$

In this case, the estimate is:

$$\hat{F}_{\infty} = \begin{pmatrix} 0.16 \\ 0.05 \\ 0.10 \\ 0.12 \\ 0.57 \end{pmatrix} . \qquad \textbf{(2.15)}$$

Element f(i) denotes the percentage of countries in quantile i.

The author concludes against (unconditional) convergence. The bimodal nature of the long run distribution points towards the division of the world into a group of haves and a group of have-nots. The prospects of transition from one group to the other are very low, as expressed in the number of zeros in the estimated Markov transition matrices.

These results are very different from the ones that spring from the application of the same method to the US states (see Quah (1994).) Here, the long run distribution is not bimodal and the series tend to concentrate in the middle quantiles.

This method was applied by Neven and Gouyette (1995) to 141 European Community regions from 1980 to 1989. They concluded in favour of a "limited poverty trap." The long run distribution was found to be more concentrated around the mean but with increased frequencies in the lower quantiles. Accordingly, the estimated transition matrix displayed a very low mobility for the poorest countries (the top diagonal element was close to 1.)

b) Markov chains and conditional convergence

Quah (1993c) provides a tentative method to test for conditional convergence using Markov chains methods. Its steps are as follows:

1) cumulate the residuals from a two-sided regression of growth on the conditioning variables. Call them g_i (t), i indexing the country.

2) Locate the absolute level of the explained growth paths by solving the following program:

$$\underset{a_1,\ldots,a_n}{Min} \sum_i \sum_t [X_i(t) - (g_i(t) + \sum_j (a_j V_j))]^2, \qquad \textbf{(2.16)}$$

X_i being the original series and V_j the time average of conditioning variables.

3) Define the transformed data that will be the subject of the tests as:

$$Y_i(t) = X_i(t) - (g_i(t) + \sum_j (a_j V_j)). \qquad \textbf{(2.17)}$$

Quah calculates two GDP per capita conditioned series. The conditioning series are physical capital, secondary school enrolment and a dummy for the African continent. These series are not used in a subsequent Markov chain estimation. Instead the author provides graphical illustrations of a *stochastic kernel* of the conditioned series' distribution. These graphs make up evidence against conditional or unconditional convergence. The statistical properties of these kernels are not known to the author, and further research in this subject is reported to be in progress at the London School of Economics.

2.3 Cointegration methods

Consider two series (Y_1 and Y_2) that have converged. From a certain point in time on, and according to the early definition of convergence (see chapter 1) , these two series are cointegrated, their cointegrating vector being (1, -1). In other words, their difference is stationary. Bernard and Durlauf (1991), Quah (1992, 1993d) and Ben-David (1994, 1995) have followed this idea.

The countries in the sample studied by Bernard and Durlauf (1991) are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden, the United Kingdom and the United States. The sample period is 1900-1987 and they are concerned with GDP per capita.

ADF tests are performed to test the pairwise cointegrating hypotheses. More than one half of the pairs cannot reject the null of any cointegration. Only six pairs of countries reject the no convergence hypothesis (i.e. that (1 -1) is not a cointegrating vector.) These are the following: France-Austria, Belgium-Holland, Denmark-Holland, France-Italy, Austria-Italy, and Finland-Germany.

The authors also use estimators based on spectral density and distribution functions. The pairwise output deviations are tested for a pure random walk using the Anderson-Darling and the Cramer - von Mises statistic. Convergence appears unlikely for all 15 countries. There are subsets of (mainly European) countries for which the convergence hypothesis cannot be rejected.

Multivariate Phillips-Ouliaris tests for cointegration are performed in three groups. Considering all the countries the authors accept the null that there are four (or five) distinct roots. For Austria, Belgium, Denmark, France, Italy and the Netherlands they accept the null that there are at least three distinct roots. When removing those six European countries from the fifteen original ones, the tests point towards the existence of five common roots.

The authors conclude that although there is a set of common factors that jointly determines international output growth, there is little evidence of convergence.

2.4 Random field methods

Quah (1992, 1993d) extends the Dickey-Fuller tests to a data set that has comparable cross-section and time-series variation (a *random field*.) Consider N series (Y_j) and T periods. A time-series cross-section regression analogous to a Dickey-Fuller regression is:

$$X_{jt} = \alpha + \beta X_{j,t-1} + u_{jt}, \tag{2.18}$$

where:

$$X_{jt} = Y_{jt} - Y_{0,t}. \tag{2.19}$$

Y_0 is a series that is chosen as a benchmark. The OLS estimator of α and β (call it a and b, respectively) obey the following properties (see Quah (1992, p. 9)):

(1) If α=0 and β=1, b converges in probability to 1 rapidly, is asymptotically normal, and the asymptotic variance can be consistently estimated.

(2) If β<1, the usual t-statistic test has asymptotic power 1;

(3) If $\alpha \neq 0$ and β=1, a and b are asymptotically normal, although converging at different rates. The covariance matrix of the estimator can be consistently estimated.

Quah (1992) contains some empirical results, considering the original series to be GDP per head for 119 economies from 1960 to 1985. Iterating over benchmarks, the author finds that income differences seem to be integrated while growth rate differences are not. Accordingly, incomes do not seem to converge, but income growth rates do.

Ben-David (1994) applied a variant of the same method to 113 countries' GDPs per head from 1960 to 1985. Here, the benchmark is the average GDP per head within the different groups considered. The author found no evidence of convergence when all the countries are included, but some evidence of convergence when the sample is restricted to the very rich or the very poor countries. In another paper, the same writer (Ben-David (1995)), using the same methodology, finds that countries that trade with each other tend to be part of the same convergence club.

This reviewer agrees with one criticism that is pointed out to these methods by Hall, Robertson and Wickens (1993). As pointed out already, cointegration is more about achieved convergence than the actual process of convergence. Two series that are converging may well have been drifting apart in the beginning of the period. In this case, cointegration tests will (rightly!) drive the researcher towards the acceptance of a no cointegration hypothesis.

The methods that use more than two variables simultaneously (multivariate cointegration and random field methods, but also some initial value and dispersion measures ones) share another drawback. If only a subset of series is converging, they may detect convergence for the whole lot, and do not provide any way of identifying the "convergence club."

2.5 The empirics of economic growth and convergence: a summary

The following points emerge from the empirical literature on convergence of incomes per head across different economies:

1) When applied to economies that are similar (US states, OECD countries, European regions, Japanese prefectures, countries that trade with each other), initial value unconditional regressions and random field regressions tend to dismiss the no convergence hypothesis. Markov chains estimation only produces results in favour of unconditional convergence in some cases[11]. Cointegration methods applied to industrialised countries do not allow for a convergence result.

[11] A convergence result for the US states is reported by Quah (1994), but Neven and Gouyette (1995) concluded against convergence across European regions using the same method.

2) When more diversified samples are considered (e. g. a large cross-section of countries), all the methods tend not to dismiss the no unconditional convergence null. Nevertheless, the negative coefficient in initial value regressions is recovered when an appropriate set of conditioning variables are added to the regression[12].

3) The significant negative coefficient in a conditional initial value regression could result from the existence of different convergence clubs as shown by Durlauf and Johnson (1992).

In chapters 3 to 7 the properties of the different methods reviewed in this chapter are evaluated. It is one of the purposes of this thesis to reconcile these different results once the properties of the different convergence tests are appraised using artificial data.

3. Kalman filter methods

Haldane and Hall (1991) and Hall, Robertson and Wickens (1992, 1993) use time varying parameters models and Kalman filter estimation methods to measure convergence. Although they apply their methods to nominal variables in the European Community, their methods are sufficiently general to be applied to other series.

According to the authors, a time varying parameters framework is particularly well suited to measure convergence. Convergence is thought to be an intrinsically structural change process and other methods would not cope with this important characteristic. For example (and as it was argued before), cointegration based methods test for achieved convergence, and not for a convergence process that is going on at the time of sampling.

Hall, Robertson and Wickens (1993) propose the following model, the parameters being estimated using the Kalman filter:

---

[12]Levine and Renelt (1992) studied the robustness of the conclusions from cross-country growth regressions to small changes in the conditioning data set. They used 119 countries from 1960-89. Although they report that almost all results are fragile, they found what they call "a qualified support for the conditional convergence hypothesis" (p. 959), in the sense that the coefficient on the initial level of income was always significantly negative when the equation included a measure of the initial level of investment in human capital.

$$X_t - Y_t = \alpha_t + \sum_{i=2}^{n} \beta_{it}(Z_{it} - Y_t) + \epsilon_t, \tag{2.20}$$

$$\alpha_t = \alpha_{t-1} + \mu_{1t}, \tag{2.21}$$

$$\beta_{it} = \beta_{i,t-1} + \mu_i, \qquad i = 2,...,n. \tag{2.22}$$

$$\epsilon_t \sim N(0, \sigma^2), \tag{2.23}$$

$$\mu_{it} \sim N(0, \Omega_{it}), \qquad i=1,...,n. \tag{2.24}$$

$$\Omega_{it} = \phi_i \Omega_{it-1}, \qquad i = 1,...,n. \tag{2.25}$$

X, Y, and the Zs are different series, and α, the βs, the Ω_{i0}s and the ϕs are parameters to be estimated.

Suppose that the ϕs have module smaller than 1. If all the βs tend to zero, then X converges to Y. If β_j tends to 1 and all the other βs tend to zero, then X converges $_j$to Z.This convergence may be said to be unconditional if α tends to 0, and is conditional otherwise.

In Hall, Robertson and Wickens (1993) this model is applied to some EC exchange, interest and inflation rates from 1972 to 1992. They considered series from Germany, the UK, France, Italy, Holland, Belgium, Spain, Ireland and Japan.[13]

The authors tested convergence of exchange rates defining X - Y as the exchange rates against the German mark and Z - Y as the German mark rate against the US dollar. The exchange rates of ERM members during the period considered were found to converge to the German mark, as opposed to the Japanese yen, the Spanish peseta and the British pound.

Convergence of interest rates was tested by making Y equal to the German rate and Z equal

[13] Japan never participated in the European Monetary System and was included as a kind of "control" series.

to the US interest rate. Although the estimated ϕs were all smaller than 1, β appeared to converge to positive values. The authors write that "the German-US differential seems to be having some effect on the differential between each EC country and Germany" (p. 12.) Note that only Belgium, Denmark, Italy, the UK and France were considered.

Inflation rates convergence was also tested by considering inflation rates differentials against Germany and the US. Hall, Robertson and Wickens (1993) found no evidence of inflation convergence.

4. VAR methods

Some recent literature on the identification of supply and demand shocks that relies on the use of vector autoregressive (VAR) methods is reviewed under this heading. Blanchard and Quah (1989) developed a theoretical framework that allowed the estimation of supply and demand shocks for the US economy. Their methodology was modified by Bayoumi (1991). Bayoumi and Eichengreen (1992a, 1992b) applied this method to US regions, the European Community and EFTA countries and Bayoumi and Sterne (1993) consider the case of 21 OECD economies.

Consider that X is a two variable vector, so that:

$$X_t = (\Delta y_t \ \ \Delta p_t)', \tag{2.26}$$

where y_t is the logarithm or real output and p_t is the logarithm of the price level.

Write X as a function of contemporaneous and lagged demand and supply shocks (ϵ_d and ϵ_s, respectively):

$$X_t = C + \sum_{i=0}^{\infty} L^i A_i \epsilon_t, \tag{2.27}$$

where $\epsilon_t = (\epsilon_{dt} \ \epsilon_{st})$ and C is a vector of constants. The identification of these shocks is achieved imposing three restrictions: demand and supply shocks are orthogonal, demand shocks have a temporary effect on output but supply ones have a permanent effect, and the variance of the shocks is equal to 1[14].

Table 2.4
Correlation of supply and demand shocks to the "central" country or region

Country or region	Supply shocks	Demand shocks
EC countries:		
Germany	1.00	1.00
Belgium	0.61	0.33
Denmark	0.59	0.39
Netherlands	0.59	0.17
France	0.54	0.35
Spain	0.31	-0.07
Italy	0.23	0.17
Portugal	0.21	0.21
Greece	0.14	0.19
United Kingdom	0.11	0.16
Ireland	-0.06	-0.08
US regions:		
Mid-East	1.00	1.00
New England	0.86	0.79
Great Lakes	0.81	0.60
Plains	0.66	0.50
Far West	0.52	0.33
South East	0.30	0.51
Rocky Mountains	0.18	-0.28
South West	-0.12	0.13

Note: The correlation coefficients refer to the entire data period: 1962-88 for EC countries and 1965-86 for the US regions.

According to the early definitions of convergence, this model implies convergence of real

[14]See Bayoumi and Sterne (1993) for more details on the estimation technique.

output if supply shocks (permanent shocks) become more and more correlated across economies and if C_1 (the deterministic trend in output) is equal across economies. Bayoumi and Eichengreen (1992a, 1992b) results can be read with this idea in mind.

Table 2.4 is reproduced from Bayoumi and Eichengreen (1992a), only the ordering of countries being different. In Europe, Germany is the central country. This role is fulfilled by the Mid-East region in the US.

From these and other results, the authors conclude that (p. 34): "a strong distinction emerges between supply shocks affecting the countries at the center of the European Community - Germany, France, Belgium, the Netherlands and Denmark - and the very different supply shocks affecting other EC members - the United Kingdom, Italy, Spain, Portugal, Ireland and Greece." In fact, the latter countries' shocks are more loosely correlated to the central ones.

Table 2.5

Percentage of variance explained by the first principal component for geographic groupings

Supply shocks	EC 11	other 11	EC core	EC periphery	Control Group	U.S. Regions
1962-88	33	26	54	32	33	49
1963-71	34	33	39	40	42	53
1972-79	44	41	63	41	51	65
1980-88	35	37	62	41	47	68
Demand shocks	**EC 11**	**other 11**	**EC core**	**EC periphery**	**Control Group**	**U.S. Regions**
1962-88	31	26	53	36	41	51
1963-71	30	34	58	30	37	44
1972-79	40	38	50	49	48	49
1980-88	40	34	54	43	56	75

Notes:
- "EC 11" means "EC countries excluding Luxembourg";
- "other 11" are other 11 non specified industrial countries;
- the "EC core" includes Germany, France, Belgium, the Netherlands and Denmark;
- the "EC periphery" includes Greece, Ireland, Italy, Spain, Portugal and the United Kingdom
- the control group comprises the US, Japan, Canada, Australia, New Zealand and Iceland.

living," and only three variables are significantly correlated with the third dimension, labelled "aviation." It would be interesting to check whether the divergence conclusion is robust in respect to other specifications for the dimensions. In particular, dimensions that account for a similar portion of the total variance could be contemplated.

Appendix to chapter 2

Initial value regressions and Galton's fallacy

Introduction

Quah (1993) argues that initial value regressions are not a very good method to test for convergence. The bulk of his argument rests on the idea that a negative cross-section coefficient on initial levels is compatible with the absence of "convergence" due to the well known Galton's fallacy of regression towards the mean.

It is shown in this appendix that Quah's (1993) argument results from the definition of convergence he adopts. Initial value regressions make some sense if the researcher is interested in other definitions of convergence but also suffer from important limitations. The latter, however, do not arise from a Galton's fallacy reasoning.

Initial value regressions and Galton's fallacy

According to one possible definition, there is convergence if the cross-section dispersion of the series declines over time. Barro and Sala-i-Martin (1995) call this "sigma-convergence" and show that "beta-convergence" does not imply it. Series "beta-converge" when differences between any two of them are stationary.

In the introduction to chapter 1 an example was provided where all series converged in the "beta" sense without necessarily implying "sigma-convergence". This was an extension of Barro and Sala-i-Martin result to a conditional convergence model.

Quah (1993) provides a closely related result: he shows that a negative coefficient in an initial value regression is compatible with a cross-section variance that does not diminish over time. Here, the bulk of his case is presented and analysed.

For the sake of argument, one can suppose that incomes in period 1 and 2 (Y_1 and Y_2, respectively) are draws from a bivariate normal variable. The expected value for Y_2

conditioned by Y_1 is given by:

$$E(Y_2|Y_1) = E(Y_2) + \lambda (Y_1 - E(Y_1)), \qquad \textbf{(2.28)}$$

where

$$\lambda = \rho_{Y_1,Y_2} \frac{\sigma_{Y_2}}{\sigma_{Y_1}}. \qquad \textbf{(2.29)}$$

ρ_{Y_1,Y_2} is the correlation coefficient between Y_1 and Y_2. If Y_1 is subtracted from both terms of equation (2.28) it gives:

$$E(Y_2 - Y_1|Y_1) = E(Y_2) - \lambda E(Y_1) + (\lambda - 1)E(Y_1). \qquad \textbf{(2.30)}$$

This is a regression of growth on the initial value. The initial value coefficient is equal to $\lambda - 1$. If the cross-section distribution is invariant, σ_{Y_1} and σ_{Y_2} being equal, this coefficient cannot be positive, because ρ_{Y_1,Y_2} is no greater than 1. Moreover, it could be negative even with time-increasing variances.

This result is related by Quah to the Galton's fallacy of regression towards the mean[19]. Galton observed that the sons of taller fathers were usually shorter than them. It would be fallacious to conclude that all men would be the same height: at the same time, the dispersion of heights was not declining and the distribution was not collapsing.

In fact, if the differences between fathers and sons were not negatively correlated to the fathers' heights the dispersion would increase over time so that people three metres or fifty centimetres tall could soon become common place. The regression towards the mean ensures convergence of heights, not in the sense that the cross-section dispersion is always declining, but in the sense that differences between any two individuals and their descendants' heights do not become increasingly bigger.

Exactly the same reasoning applies to countries' incomes. If countries' incomes regress

[19]See also Maddala (1992) for a concise exposition of the Galton's fallacy. Friedman (1992) also criticises initial value regressions following a reasoning that is close to Quah's.

towards the mean, this is a sign of convergence, and not the contrary; it is not possible to have a country that becomes richer and richer over time compared to one that gets poorer and poorer. The differences between any two countries' incomes are therefore stationary.

If g is the income growth rate ($g=Y_2-Y_1$), it is easily shown that:

$$\lambda = 1 + \rho_{g,Y_1} \frac{\sigma_g}{\sigma_{Y_1}}. \qquad \textbf{(2.31)}$$

A negative coefficient only arises if there is a negative correlation between growth rates and initial values. It can be safely said that if there is no correlation between growth rates and initial values, the coefficient equals zero, there is no regression towards the mean and consequently there is no convergence. This result could be considered to be evidence in favour of endogenous growth models;. in those, growth rates are not a function of initial values and usually depends on country-specific factors, like policies.

However, the converse is not true. A negative correlation between growth rates and initial incomes does not imply that countries are all regressing towards the mean (and therefore converging in the “beta” sense). Namely, they may well be regressing towards two or more different means (more than one convergence club), as it was shown by Durlauf and Johnson (1992). Alternatively, it could happen that only a subset of incomes' series is converging to the same mean, the others not converging at all (limited convergence). These two possibilities allow for a negative ρ_{g,Y_1} in equation (2.31) and are studied in chapters 5 and 6, respectively. The main idea conveyed here is that a statistically negative coefficient on the initial value is not sufficient to establish a convergence result in the “beta” sense. This does not arise because of the Galton fallacy and the negative correlation between growth rates and initial conditions is still an interesting finding that ought to be explained.

Chapter 3
Unconditional Convergence

Introduction

The empirical literature on convergence relies on the use of different methods for testing convergence of economic series. Some methods are not real or complete tests, in the sense that there are no involved confidence levels. Other methods imply a detailed testing procedure. In this chapter, a test for convergence is a test where the null hypothesis is the no convergence one. Therefore, the burden of proof is put on the convergence side. This is not the way convergence is assessed in part of the literature but some normalisation was needed in order to compare different approaches. Hopefully, some welcome additional rigour is provided with this formulation.

The absolute and comparative advantages of these methods are little known. Namely, the power of each test under different alternatives has not been subject to much scrutiny. In this chapter, the power of several tests is assessed when the alternative hypothesis implies the convergence of a whole set of series.

The general approach was the following:

1st step - To simulate several replications of artificial data that are converging. Different "speeds of convergence" and different "initial distances" are considered (when applicable.)

2nd step - To apply different convergence tests to each replication. To count the number of times the null of no convergence was rejected, i.e., to compute the power of each test.

3rd step - To assess the dependence of "power" on the "speed of convergence" and on the "initial distance."

4th step - To compare the power of different methods under similar circumstances.

The first part of the chapter describes the data generation process. In all cases, all the considered series converge unconditionally to the same steady-state. The meaning of "speed of convergence" and "initial distance" is clarified.

The different methods or tests are presented in the second part. These are methods that rely on dispersion measures, nonparametric methods, initial value regressions, Dickey-Fuller tests for cointegration, random field regressions, the Markov chains, and the Kalman filter method.

The results are discussed individually for each method in the third part and compared in the fourth part. The chapter ends with the conclusion.

1. The data generation process

The different methods were applied to a set of 100 artificially generated series. Each series consists of 100 time observations, except otherwise indicated. In the following lines, x_{it} is the value of series i in period t. i and t may assume any integer value between 1 and 100. X_i is the vector of 100 observations for the i-th series.

X_1 is the attracting series. It was generated according to the following mechanism:

$$x_{1,t}=x_{1,t-1}+g+\eta_{1,t}+\Delta\epsilon_{1,t}, \tag{3.1}$$

where g is a constant and η and ϵ are mean zero normally distributed uncorrelated random variables with no autocorrelation.

The 99 attracted variables were generated obeying the following equation:

$$x_{i,t}=x_{i,t-1}+g+\beta(x_{1,t-1}-x_{i,t-1})+\eta_{i,t}+\Delta\epsilon_{i,t}, \qquad i=2,3,\ldots,100. \tag{3.2}$$

β (the speed of convergence) is non-negative and smaller than 1. Again, η and ϵ are mean zero normally distributed uncorrelated random variables.

Irrespective of the initial gap between series 1 and i, their difference becomes equal to a mean zero stationary random variable in the long run (when t tends to infinity.) Accordingly, the series converge unconditionally. This is according to one possible definition of converging series[1].

To generate the wanted 100x100 artificial data matrix one not only needs equations (3.1) and (3.2) but also to define a set of initial values (one for each series), the values for the variances of the random variables, and the value for the constant g.

The initial values were the natural logs of the GDP per head using international prices in a set of 100 countries in 1960. The chosen countries were the more populated ones among the 113 countries for which there was complete data from 1960 to 1989.

The raw data for the initial values are shown in the appendix. The data came from the Penn World Table (Mark 5.5), available through Internet from the NBER gopher in the US. This data set is a revised and updated version of the Mark 5, described in Summers and Heston (1991) and was often used in the convergence empiric literature.

g was made equal to the US average growth rate between 1960 and 1989. The random variables η and ϵ represent permanent and transitory shocks to the series, respectively. Their variance was considered to be equal. The total variance of the growth rate was made equal to the sample variance of the US growth rate between 1960 and 1989.

Accordingly:

$$Var(\epsilon) = Var(\eta) = \frac{1}{3} Var(US\ growth\ rate), \qquad \textbf{(3.3)}$$

with:

$$Var(US growth rate) = 0.00070107, \qquad \textbf{(3.4)}$$

and

$$g = 0.0217. \qquad \textbf{(3.5)}$$

[1]See the introduction to chapter 1 for the definition of unconditional convergence.

2. The methods

2.1 Dispersion measures

Define V, the sample variance, as:

$$V_t = \sum_i \frac{(x_{i,t} - \bar{x}_t)^2}{99}. \qquad \textbf{(3.6)}$$

A declining pattern of this dispersion measures is usually interpreted as evidence of convergence[2].

Under the null hypothesis of equality of variances in the first and last period, the following ratio follows an F(98, 98) distribution[3]:

$$\frac{V_{100}}{V_1} \frown F(98,98). \qquad \textbf{(3.7)}$$

A ratio that is big enough leads to the dismissal of the no convergence null. (The 5 and 1 percent critical values for the relevant F distribution are respectively 1.394 and 1.6.)

Note that the "no convergence null hypothesis" is defined here in terms of a dispersion measure. It was shown in Chapter 2 that an increasing variance is compatible with other definitions of convergence. It is possible for instance to have a positive β and an increasing time path for the cross-sectional variance, given a sufficiently tight initial distribution.

---

[2]Several authors have used dispersion measures methods when assessing convergence of economic series. See Baumol and Wolf (1988), Lichtenberg (1994) and Barro and Sala-i-Martin (1995), among others, for an application to GNP per capita. Tovias (1982) and Mohktari and Rassekh (1989) use these techniques when testing for factor prices convergence across countries. Chapter 2 reviews this literature.

[3]See Lichtenberg (1994) for more details.

2.2 Nonparametric methods

Two nonparametric tests are applied to the convergence framework. These are the *sign test* and the *Page's test for ordered alternatives*[4].

In both cases a distance from the mean is defined. Let $m_t = \sum_i \frac{x_{it}}{n}$ be the cross-section mean. Three possible distances are:

$$d^1_{i,t} = |x_{i,t} - m_t|, \quad \textbf{(3.8)}$$

$$d^2_{i,t} = (x_{i,t} - m_t)^2, \quad \textbf{(3.9)}$$

$$d^3_{i,t} = |\frac{x_{i,t}}{m_t} - 1|. \quad \textbf{(3.10)}$$

The sign test

This test can be used in the comparison between any two periods. Suppose that we are comparing period S to period T (S<T.)

Let $D_i = d_{iT} - d_{iS}$. M_D is the median of D.

The null hypothesis is:

H_o: $M_D = 0$ (there is no convergence between S and T),

against

H_a: $M_D < 0$.

The test procedure is as follows:

[4]A description of both these tests in general terms can be found in Daniel (1990). As far as this author knows, they have not yet been used in any empirical work on convergence.

Record a "+" if $D_i>0$ and a "-" if $D_i<0$ and eliminate the observation if $D_i=0$. Let S be the number of "+" signs. If there is convergence between periods S and T, one would expect $d_{is} > d_{it}$. Consequently, a sufficiently small number of "+" leads us to reject H_0[5].

The critical value S_α is obtained using the following large-sample approximation:

$$s_\alpha = 0.5n + z_\alpha\sqrt{0.25n}, \quad \textbf{(3.11)}$$

where n is the total number of "+" and "-" signs and z_α is the negative standard normal variable value that corresponds to the α significance level.

In the implementation of this test the first and last period were considered, so that S=1 and T=100.

<u>Page's test for ordered alternatives</u>

This test allows us to compare more than two periods. Supposing that there are k ordered periods to compare, the null hypothesis of no convergence becomes:

H_0: $M_{D,1} = M_{D,2} = \ldots = M_{D,k}$ (no convergence),

and the alternative is:

H_a: $M_{D,1} > M_{D,2} > \ldots > M_{D,k}$ (convergence),

where $M_{D,i}$ is the median of the distances in period k.

The test statistic procedure is as follows:

Rank each distance series according to period and compute the sums of ranks R_j for each period.

[5]This is a binomial test, where the null of no convergence is H_0: $p_+=0.5$, p_+ being the proportion of "+." See Daniel (1990) for more details.

Compute:

$$L = \sum_j jR_j. \tag{3.12}$$

For large samples[6], the following statistic is approximately standard normal under the null:

$$z = \frac{L-(Nk(k+1)^2/4)}{\sqrt{N(k^3-k)^2/144(k-1)}}. \tag{3.13}$$

This last statistic was therefore used as part of a nonparametric convergence test. The periods considered were periods 1, 20, 40, 60, 80 and 100.

2.3 Initial value regressions

This test for convergence is probably the more commonly used throughout the empirical literature on convergence of incomes per head[7]. The Barro and Sala-i-Martin (1992a) version consists in the estimation of the following cross-section equation[8]:

$$\frac{x_{i,T}-x_{i,1}}{T-1} = constant + \frac{(1-\lambda)^{T-1}-1}{T-1}x_{i,1} + v_{i,T}. \tag{3.14}$$

The average growth rate for each series is regressed on the series initial value. This is a test for *unconditional* convergence, λ being the speed of convergence. Equation (3.14) can be derived from the neoclassical exogenous growth model. A statistically negative estimated coefficient on the initial value (meaning a positive speed of convergence) is taken as evidence of convergence.

Equation (3.14) is estimated using the ordinary least squares (OLS) method. The test for

---

[6]For small samples, see Daniel (1990) for tables.

[7]See, among others: Barro and Sala-i-Martin (1992a, 1992b, 1995), Baumol (1986), Dowrick and Nguyen (1989), Dowrick (1992), Durlauf and Johnson (1992), Knight, Loyasa and Villanueva (1993) and Mankiw, Romer and Weil (1992). Chapter 2 presents and discusses empirical methods and results based on initial value regressions.

[8]Note that $(1-\lambda)^T \approx e^{-\lambda T}$, for sufficiently small values of λ. Barro and Sala-i-Martin actually use the "e" version.

convergence involves the calculation of the t-statistic on the initial value coefficient (λ is estimated from this coefficient.)

Under the null of no convergence, this statistic should asymptotically follow a standard normal distribution, so that negative values imply that convergence is taking place.

The data generation mechanism as expressed by equations (3.1) and (3.2) implies the following equation:

$$\frac{x_{i,T}-x_{i,1}}{T-1} = \frac{x_{1,T}-(1-\beta)^{T-1}x_{1,1}}{T-1} + \frac{(1-\beta)^{T-1}-1}{T-1}x_{i,1} + (T-1)^{-1}\sum_{t=2}^{T}(1-\beta)^{T-t}u_{i,t} \quad \textbf{(3.15)}$$

with

$$u_{i,t} = \eta_{1,t} - \eta_{i,t} + \Delta\epsilon_{1,t} - \Delta\epsilon_{i,t}. \quad \textbf{(3.16)}$$

When (3.15) is compared to (3.14), one concludes that they are actually the same, with $\beta=\lambda$. Therefore, the estimates of λ are estimates of β, which is a pre-defined parameter in the Monte Carlo experiments.

2.4 Dickey-Fuller tests

These tests are based in the idea that if two series converge their difference should be stationary[9]. This is not always true: two series can converge even if they are not cointegrated. To understand this more clearly, it is enough to imagine the extreme case of two series that are independent random walks until a certain point in time, and that from then on are exactly equal.

With the DGP described above, two converging series are indeed cointegrated, their difference being stationary. Subtract any of the series from the first (attracting) series:

[9]This was the approach of Bernard and Durlauf (1991), who applied cointegration Dickey-Fuller tests, among other cointegration tests, to test for GDP per capita convergence across countries. See Chapter 2 for a review of their findings.

$$x_{1,t} - x_{i,t} = d_{i,t} = (1-\beta)d_{i,t-1} + u_{i,t}, \qquad i=2,...,N. \tag{3.17}$$

where u is defined by (3.16).

Since β is smaller than 1 but non-negative, and u is a stationary and normally distributed random variable, the d series is stationary.

The Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests imply the estimation of equations like the following[10]:

$$d_{i,t} - d_{i,t-1} = c + k_0 d_{i,t-1} + k_1 \Delta d_{i,t-1} + ... + k_m \Delta d_{i,t-m} + w_t. \tag{3.18}$$

In the DF-no constant case, c=0 and $k_1 = ... = k_m = 0$ are imposed from the beginning.

If c is allowed to be different from 0, but $k_1 = ... = k_m = 0$, we have the "DF with a constant" test.

The ADF(m) test implies that the parameters k_1 to k_m are not constrained to be 0.

In any case, the conventional OLS t-statistic on the k_0 parameter is calculated. Under the null hypothesis of non-stationarity, its distribution is non-standard, and some "Dickey-Fuller" tables have to be consulted.

Usually, the the number of lags given by m in equation (3.18) is chosen so that the error term does not show evidence of autocorrelation.

2.5 Random field regressions

A random field is a data set where the time-series and cross-section dimensions have comparable magnitudes. Quah (1992, 1993d) extends the unit-root regression to a random field framework. The generated data is a random field: the time-series dimension is exactly

[10]They can as well include a time trend term. This case was not considered.

equal to the cross-section one, since we have 100 periods and 100 series.

Following Quah (1993d), let b_T be the estimator for the regression coefficient of $d_{i,t}$ on its own first lag:

$$b_t = (\sum_{j=2}^{W} \sum_{t=1}^{T} d_{j,t-1}^2)^{-1}(\sum_{j=2}^{W} \sum_{t=1}^{T} d_{j,t}\ d_{j,t-1}). \qquad \textbf{(3.19)}$$

If d has a unit root, so that $\Delta d_{i,t}=\mu_{i,t}$, it can be shown that, under a set of assumptions[11]:

$$z_f = T\sqrt{\frac{N}{2}}(b_T-1) \rightarrow N(0,1) \qquad as\ T\rightarrow\infty. \qquad \textbf{(3.20)}$$

This statistic was therefore used to test for the stationarity of the differences between each series and a benchmark series. The chosen benchmark was the first series, as it was for the Dickey-Fuller tests.

2.6 Markov chains method

The initial cross-section distribution of the series (incomes) is quite unequal. If there is any convergence among the series, it is expected that this distribution will become somehow more equal. Moreover, the initial position of a series should not matter in the long run: the lowest series in the first period and the highest one should have the same probability of being at the top in the distant future.

Suppose we define a grid of values in the form of a vector:

$$H = \begin{pmatrix} 0.25 \\ 0.50 \\ 1.00 \\ 2.00 \end{pmatrix}. \qquad \textbf{(3.21)}$$

The grid H implicitly describes the cross-sectional distribution of the new series into 5 quantiles. Considering that the distribution is given by a five-element vector F_t, its first element

[11]See Quah (1993d) for more details.

should equal the proportion of series that are smaller than a quarter of the average at time t, its second element should be equal to the proportion of series that fall between a quarter and a half of the average, and so forth; finally, its fifth and last element will be identical to the fraction of the series that exceed two averages.

The elements and size of H could be different. Those were chosen because they divide roughly the initial values in roughly equally sized groups and also because they were previously used by the proponent of this method, Quah (1993a), in an empirical study that used a similar data set.

Assuming now that the law of motion of the distribution is described by a matrix M:

$$F_{t+1} = M.F_t. \qquad \textbf{(3.22)}$$

M is a time invariant Markov transition matrix. The m_{ij} element denotes the probability of a series that is in the j-th quantile in time t to swap into the i-th quantile in time t+1.

The maximum likelihood estimation is achieved by averaging the actual transition proportions through time[12].

An estimate of the long run distribution of the series is given by:

$$\hat{F}_\infty = \lim_{s\to\infty} \hat{M}^s \cdot F_t, \qquad \textbf{(3.23)}$$

where $\hat{M}$ is an estimate of the transition matrix.

There will be evidence of convergence if the following conditions are fulfilled:

1) The non-diagonal elements of $\hat{M}$ are significantly positive, so that a series is not locked in its initial quantile.

2)The long run distribution is such that the extreme quantiles are virtually empty and the series

[12]See Lee, Judge and Zellner (1970) for a proof.

concentrate into the central ones.

Two transition matrices were estimated. In the first case, a one-period transition matrix was considered. The second estimation produces a 100-period transition matrix. The associated long run distribution was computed in both cases.

2.7 Kalman filter method

One version of a convergence test proposed by Hall, Robertson and Wickens (1993) was implemented. Consider the following formulation, where i is a natural number between 2 and 100:

$$x_{1,t} - x_{i,t} = \alpha_{i,t} + \epsilon_{i,t}, \tag{3.24}$$

$$\alpha_{i,t} = \alpha_{i,t-1} + \mu_{i,t}, \tag{3.25}$$

$$\epsilon_{i,t} \sim N(0, \sigma_i^2), \tag{3.26}$$

$$\mu_{i,t} \sim N(0, \Omega_{i,t}), \tag{3.27}$$

$$\Omega_{i,t} = \phi_i^2 \Omega_{i,t-1}, \tag{3.28}$$

$$\Omega_{i,0} = \Upsilon^2. \tag{3.29}$$

The difference between each series and the attracting series is supposed to be a random walk plus noise. The noise variance (σ^2_i) is considered to be invariant through time. Not so for the variance of μ_i. This variance may display a declining pattern: if ϕ_i is less than 1, it tends to 0 in the long run. This means that the two series x_1 and x_i are converging, their difference becoming a stationary variable.

Table 3.1
Critical values for the $T(\phi_{i,ML})$ statistic

%	t-value
0.5	-3.702
1	-3.479
5	-2.479
10	-1.970
50	-0.059
99	3.348
99.5	3.529

The model described by equations (3.24) to (3.29) is written in state-space form. Equation (3.24) is the measurement equation and equation (3.25) is the state one. Therefore, the likelihood function can be built using the Kalman filter. Maximum likelihood (ML) estimates for the parameters are obtained by maximising it[13].

Under the null of no convergence, ϕ_i=1. Therefore, the proper test for convergence is: H_0: ϕ_i=1 against H_a: ϕ_i<1. To actually implement this test, one needs to know the distribution of (some function of) ϕ_i under the null. After 1000 replications, the distribution of the following was tabulated:

$$T(\phi_{i,ML}) = \frac{\phi_{i,ML} - 1}{\sqrt{(h^{-1})_{22}}}. \qquad \textbf{(3.30)}$$

[13]References on the Kalman filter are Harvey (1989) and Cuthbertson, Hall and Taylor (1992). See also Hall, Robertson and Wickens (1992).

In (3.30), $(h^{-1})_{22}$ is the second element in the diagonal of the inverse of the Hessian matrix. It is common to take this as an estimate of the variance of the second ML parameter[14]. $T(\phi_{i,ML})$, although not very different from a normal distribution, is more dispersed than the standard normal. It appears that the Hessian estimator is underestimating the true variance, as already mentioned by Hall, Robertson and Wickens (1993) . The tabulation is reproduced in Table 3.1

.

2.8 A first comparison of methods

The eight different methods can be divided into two groups. The first group comprises all the methods that deal with the whole cross-section of data (they can be called cross-section methods), the second group including those methods that use two series only (called time-series methods.) Consequently, the Dickey-Fuller and Kalman filter methods belong to the second group, all the other methods falling in the first category.

This is a crucial difference. The null hypothesis of no convergence across *all* the series implies that none of the series converges to another. On the other hand, the null hypothesis of no convergence of *one* series to another does not imply that the other series are not converging, or even that any of the series being considered are not converging to a third one.

An important issue arises from these considerations: what if the true data generation process forces limited convergence, in the sense that only a subset of series converge? This issue is not treated in this chapter, but it is very possible that methods that use all the series could lead the researcher into either not to reject the no convergence hypothesis or to reject the null of no convergence in favour of an (erroneous) alternative of complete convergence across series. An important research purpose can well be to separate the series between those that are converging and the ones that are not[15]. This task is not easy to achieve using the methods included in the first group, most or all of them requiring some modifications or extensions that have yet to prove their effectiveness.

As explained later in this chapter, these considerations are the motivation for further research

[14]The other parameter is the initial standard deviation, Υ.

[15]As in Hall, Robertson and Wickens (1992, 1993).

on the methodology of convergence measurement.

3. The results

3.1 Dispersion measures

Results from 1000 replications for two different values for the speed of convergence are presented in Table 3.2.

Table 3.2
Ratio of variances (V_1 / V_{100})
(1000 replications)

β value: V_1 / V_{100}	0.00	0.02
Minimum	0.860	28.8
Maximum	1.08	59.2
Average	0.972	38.8

The first column displays the values that resulted from 1000 replications when the null hypothesis of no convergence was the true one, i. e. , when the speed of convergence (β) is equal to zero. In that case, the distribution of the ratio of the variances is not centred in 1, as could be expected. This happened because the data generation process included permanent uncorrelated shocks to each series, which made the theoretical dispersion across series increases through time. A test for the equality of variances using the F-statistic and critical values presented in section 2.1 never leads to the rejection of the null hypothesis in favour of an alternative of V_{100} being smaller than V_1.

When the speed of convergence equals 0.02, the ratio is always much higher than one, meaning that the variance in period 100 is much smaller than the initial variance. Accordingly, the F test for the equality of variances leads to the rejection of the null hypothesis in all cases. Experiments with higher speeds of convergence produced even more unequal variances and are not reproduced in the table.

3.2 Nonparametric methods

When applying nonparametric methods, the first formulation of the distance from the mean was used:

$$d_{i,t}^{1}=|x_{i,t}-m_t|. \tag{3.31}$$

If one thinks of the generated series as GDPs per head in log form, a natural interpretation considering the data generation process and the starting points, this distance is close to the percentage gap between a country GDP and the average GDP.

The sign test and Page's test were applied to 1000 replications for each of the three different speeds of convergence and also for a no convergence situation (β=0.00.)

Table 3.3
Nonparametric methods
Percentage of rejections of the no convergence hypothesis
(1000 replications)

β value:	0.00	0.02	0.10	0.50
sign test 5% significance (1%)	1.3 (0.2)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Page's test 5% significance (1%)	0.0 (0.0)	100.0 (100.0)	100.0 (76.5)	96.3 (0.1)

When the null of no convergence is true (first column in Table 3.3), this hypothesis was rejected only a few times. As in the dispersion measures method, this happened because the "no convergence" data generation process implies an increasing (and not a constant) dispersion through time.

Consider now the positive values of β, and first the sign test. In every replication, irrespective of the value of β, the no convergence null hypothesis was rejected, even at the 1%

significance level. This test relies on the comparison of dispersion in the last period with dispersion in the first period. One would expect that the power of this test power would increase (or at least remain the same) when the speed of convergence gets higher.

The results are different when one considers the Page's test. Its power seems to decrease with β. When β=0.50, one can almost never reject the no convergence hypothesis at the 1% level. This should not be surprising if one considers the nature of the null and alternative hypothesis. As stated before, the no convergence hypothesis implies that dispersion is the same in periods 1, 20, 40, 60, 80 and 100. Under the alternative convergence hypothesis, one assumes that dispersion is declining from period to period. If convergence is very strong (if β is high), it may happen that convergence is achieved before period 100. If, say, most of the convergence process occurs before period 20, the test would not be very accurate, because dispersion would actually be equal from then on.

3.3 Initial value regressions

Equation (3.14) was estimated 1000 times for each of the three different speeds of convergence. Results obtained when the null of no convergence is the true hypothesis are also included. λ_{LS} is the estimate of the speed of convergence derived from the OLS estimate of the coefficient on the initial value. More precisely:

$$\lambda_{LS} = 1 - [(T-1).C+1]^{\frac{1}{T-1}}, \tag{3.32}$$

where C is the estimated coefficient. If C is sufficiently negative so that (T-1).C is less than 1, one can not estimate the speed of convergence. This was the case in almost half of the times when the true speed was 0.10 or 0.50 (the percentage of times it was possible to infer λ_{LS} is denoted in the table as "percentage of successes.")

λ_{LS} was a good estimate of the speed of convergence only when β was sufficiently low (less or equal than 0.02) , as shown in Table 3.4.

Table 3.4
Initial value regressions
(1000 replications)

β value:	0.00	0.02	0.10	0.50
average λ_{LS}	0.0000	0.0200	0.0604	0.0641
min λ_{LS}	-6.23×10^{-4}	0.0182	0.0434	0.0457
max λ_{LS}	5.38×10^{-4}	0.0223	0.113	0.124
average t-statistic	2.35×10^{-3}	-98.1	-231.	-354.
min t-statistic	-3.23	-123.	-295.	-448.
max t-statistic	3.54	-71.3	-185.	-283.
% of non convergence rejections 5% (1 %)	4.8 (1.4)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
% of successes	100.0	100.0	49.3	49.6

Nevertheless, the null of no convergence was always rejected, even at the 1% significance level and when the positive value of β was smaller (the maximum value for the t-statistic was -71.3 for this set of replications, well below its critical value.)

On the whole, this was a very powerful test. This is not surprising, since the data generation mechanism implied that all the series were converging at the same speed.

3.4 Dickey-Fuller tests

The differences between the attracting series and series 2, 10, 50 and 100 were tested for stationarity. Again, 1000 replications were calculated under the three different speeds of convergence. The critical values, presented in Table 3.5, were taken from MacKinnon (1991).

Table 3.5
ADF tests
critical values

1 percent	5 percent
-3.497	-2.890

The ADF(2) statistic was always used except in three cases: when the speed of convergence was equal to 0.50 with series number 50 and 100, and with speed equal to 0.10 with series 100. In these cases, an additional lag was needed to induce non-autocorrelated residuals, so that ADF(3) was computed.

Table 3.6
Augmented Dickey-Fuller tests
T=100
Percentage of no convergence rejections
(1000 replications)
5 % confidence level
(1 %)

initial position: / speed of convergence:	2	10	50	100
0.00	6.4 (1.5)	5.4 (1.1)	6.3 (1.2)	6.9 (1.3)
0.02	9.2 (2.1)	8.3 (1.9)	87.9 (63.6)	99.9 (99.9)
0.10	38.8 (13.2)	99.1 (88.3)	100.0 (100.0)	100.0 (100.0)
0.50	99.9 (99.7)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)

In Table 3.6, results are shown for each of the four different speeds of convergence including the zero speed (no convergence) case. Each case contains the percentage of times the t-statistic (defined in section 2.4) was lower than the five and one percent critical levels, leading to the rejection of the no convergence hypothesis. When $\beta = 0.00$, the ADF statistic seems to follow the Dickey-Fuller distribution, as expected from the data generation process.

Results presented in Table 3.6 correspond to a number of observations equal to 100. In some real world situations, the number of periods available is smaller than this number. Since the power of time series methods usually depend on the number of periods (given by T), it was felt that there was the need for another experiment with a smaller number of observations. These results are presented in Table 3.7, where T equals 40. This is roughly the number of years that growth studies can use when assessing convergence after the Second World War. It is apparent from a comparison between Tables 3.6 and 3.7 that the power of the test decreases when T is smaller.

Table 3.7
Augmented Dickey-Fuller tests
T=40
Percentage of no convergence rejections
(1000 replications)
5 % confidence level
(1 %)

initial position: / speed of convergence:	2	10	50	100
0.02	7.3 1.3	3.8 0.5	9.0 2.0	36.6 14.6
0.10	10.4 2.9	56.3 24.6	99.9 98.0	97.3 81.5

Three main conclusions can be derived from results displayed in Tables 3.6 and 3.7:

- The tests are more powerful when β is higher. This is another version of the otherwise well known result that unit root tests are not very powerful when the root is close to 1.

- The tests are more powerful when the initial difference is higher. For instance, when the difference between series 1 and 100 was considered, the null of no convergence was rejected

99.9 percent of the times at the 1 percent significance level, even when $\beta = 0.02$.

- The tests are more powerful when the number of available observations is higher.

3.5 Random field regressions

Table 3.8
Random field regressions
(1000 replications)

β value:	0.00	0.02	0.10	0.50
average b_{OLS}	1.000	0.979	0.898	0.491
min b_{OLS}	0.998	0.975	0.887	0.466
max b_{OLS}	1.002	0.983	0.909	0.521
min z_f	-1.682	-17.75	-79.77	-377.7
max z_f	1.656	-11.91	-64.28	-338.9
% of no convergence rejections 5% (1%)	0.2 (0.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)

As in the initial value regressions, the estimates of b proved to be good estimates of the speed of convergence. In 1000 replications, the average of b was always close to $1-\beta$, as shown in Table 3.8.

For each of the three values of β, the no convergence null was always rejected, even at the 1% level. Under the null, the statistic z_f would asymptotically follow a standard normal distribution. In the replications the distribution of z_f was tighter than the standard normal. This is most probably due to the fact that the considered DGP implies autocorrelation in the error term, and therefore does not obey one of the conditions specified by Quah (1993d). When $\beta=0.02$, its maximum value was -11.91, quite far from the critical value. On the whole, the null of no convergence was rightly rejected every time.

3.6 Markov chains method

This method was applied to a situation where the speed of convergence is low (β=0.02.) The results were so overwhelmingly in favour of convergence that no other experiments were made with higher speeds, the nature of the method implying that the convergence evidence would be comparably strong.

1000 replications were computed. Results were very similar across iterations, so the results of the first one are presented as an example.

The estimated 100-period transition matrix was:

$$M_{100} = \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.833 & 0.794 & 0.583 & 0.417 & 0.111 \\ 0.167 & 0.206 & 0.417 & 0.583 & 0.889 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix} \quad \textbf{(3.33)}$$

The (i,j) element of this matrix is the percentage of series in j-th quantile in period 1 that swapped into the i-th quantile in period 100. One immediately notices that the first, second and fifth quantiles were empty in the final period.

The corresponding long run transition matrix is equal to:

$$\lim_{t\to\infty} M_{100}^{t} = \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix} \quad \textbf{(3.34)}$$

The probability that a series will belong to the central third or fourth quantiles in the long run is always equal to 0.5. This does not depend on the initial position of the series. Therefore, the associated long run distribution is:

$$F_\infty = \begin{pmatrix} 0.0 \\ 0.0 \\ 0.5 \\ 0.5 \\ 0.0 \end{pmatrix} . \tag{3.35}$$

Series are concentrated in the middle quantiles; convergence is at work.

Similar results arise when a one-period transition matrix is considered. The estimated matrix for this same replication was:

$$M_1 = \begin{pmatrix} 0.820 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.180 & 0.894 & 0.003 & 0.0 & 0.0 \\ 0.0 & 0.106 & 0.973 & 0.032 & 0.0 \\ 0.0 & 0.0 & 0.024 & 0.966 & 0.098 \\ 0.0 & 0.0 & 0.0 & 0.002 & 0.902 \end{pmatrix} \tag{3.36}$$

Although the diagonal elements are the highest, the probability of a series to swap into a neighbouring quantile, and specially to the one closer to the middle point, is not at all negligible. For example, the reader may notice that a series that is in the lowest quantile has almost a 20% probability of changing into the second lowest quantile next period.

The long run transition matrix derived from the one-period one is:

$$\lim_{t\to\infty} M_1^t = \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.018 & 0.018 & 0.018 & 0.018 & 0.018 \\ 0.553 & 0.553 & 0.553 & 0.553 & 0.553 \\ 0.418 & 0.418 & 0.418 & 0.418 & 0.418 \\ 0.011 & 0.011 & 0.011 & 0.011 & 0.011 \end{pmatrix} \tag{3.37}$$

Although not equal, this matrix shares the basic properties of the matrix in expression (3.34): the probability of a series belonging to the lowest or highest quantiles is 0, or nearly 0, in the long run, and these probabilities do not depend on the initial position.

Again, the long run distribution is concentrated in the middle quantiles, and this fact is taken into account as "convergence evidence":

$$F_{\infty} = \begin{pmatrix} 0.0 \\ 0.018 \\ 0.553 \\ 0.418 \\ 0.011 \end{pmatrix} . \qquad \textbf{(3.38)}$$

3.7 Kalman filter method

The Kalman filter method is applied to two series at a time. This peculiarity is one of its strengths: there is at least the hope that this method will identify the subset of the series that are converging, in a case of limited convergence.

Table 3.9
Kalman filter estimations
T=100
β=0.02
(1000 replications)

series:	2	10	50	100
average ϕ_{ML}	1.000	0.998	0.995	0.993
max ϕ_{ML}	1.025	1.017	1.005	1.000
min ϕ_{ML}	0.980	0.968	0.980	0.984
% of no convergence rejections 5% (1%)	3.9 (0.4)	7.6 (1.7)	32.6 (9.4)	72.4 (39.0)

The data generation mechanism generates 100 series that converge to each other, because they are all attracted by the first one. The series are ordered according to their initial distance from the attractor, and the power of the Kalman filter method was assessed in different initial

Table 3.10
Kalman filter estimations
T=100
β=0.10
(1000 replications)

series:	2	10	50	100
average ϕ_{ML}	1.000	0.991	0.976	0.965
max ϕ_{ML}	1.039	1.006	0.994	0.985
min ϕ_{ML}	0.959	0.928	0.930	0.913
% of no convergence rejections 5% (1%)	5.9 (1.0)	48.9 (21.3)	99.4 (97.9)	100.0 (99.5)

conditions. Different speeds of convergence were also imposed. Specifically, 1000 replications for series 2, 10, 50 and 100[16] and β values equal to 0.02, 0.10 and 0.5 were considered.

It must be beared in mind that a ϕ_{ML} statistically smaller than 1 implies a rejection of the null of no convergence. The 1% and 5% critical values are (see Table 3.1), respectively, -3.479

Table 3.11
Kalman filter estimations
T=100
β=0.50
(1000 replications)

series:	2	10	50	100
average ϕ_{ML}	0.946	0.809	0.656	0.595
max ϕ_{ML}	1.071	1.005	0.985	0.953
min ϕ_{ML}	0.000	0.062	0.364	0.438
% of no convergence rejections 5% (1%)	8.3 (2.5)	87.2 (57.9)	99.5 (98.3)	99.6 (99.4)

[16]As a curiosity, these series correspond respectively to Switzerland, Venezuela, Malaysia and Lesotho. Series 1 is the US.

and -2.479, the statistic being the one expressed in (3.30).

The main results are reproduced in Tables 3.9 to 3.11.

It is very clear that the power of the test increases with the speed of convergence and with the initial gap between the series.

Consider series 2. It starts very closely to the attractor. In a sense, one can say that this is a series that has already converged, as opposed to series that are still converging. Even when β=0.50, the percentage of rejections is not much higher than the significance level. Moreover, the average ϕ_{ML} is equal to 1. In this case, one cannot distinguish between two converging series and two series that display no convergence at all (i.e. series for which β=0.00.)

Table 3.12
Kalman filter tests
T=40
Percentage of no convergence rejections
(1000 replications)
5 % confidence level
(1 %)

initial position: / speed of convergence:	2	10	50	100
0.02	5.7 1.1	7.6 1.7	10.4 1.8	20.1 1.4
0.10	5.9 1.1	28.2 6.6	99.3 97.7	100.0 99.3

At the other extreme, one can think of the results for series 100. Even when the speed of convergence is low (β=0.02), one rejects the no convergence hypothesis at the 5% significance level almost three in four times. This proportion grows to very close to 1 for higher speeds.

As with the other time series method, power declines with the number of observations. Table

3.12 includes results for two speeds of convergence (0.02 and 0.10) when T is equal to 40.

4. Results compared

For the sake of analysis, the methods can be divided into two groups. The first group comprises the methods that use the whole cross-section of series. The second group includes the methods that assess convergence of only one series to another. The Kalman filter and the Dickey-Fuller methods are part of the second group. All the other methods belong to the first one.

The methods in the first group proved to be very powerful. For some of them (initial value and random field regressions) there was not a single time when the no convergence hypothesis was not rejected. Only the Page test seemed to lose power when the speed of convergence was equal to 0.5. Even in this case, this is most probably due to the specific formulation of the null hypothesis.

These results arise because of two important and restrictive characteristics of the data generation process: every series is converging to the first one, and at the same speed. It comes as no surprise that either the random field or the initial value regressions capture this special pattern of convergence, being able to correctly estimate the common speed. But what if the speeds were different and/or some of them were equal to zero? It could well happen that the no convergence hypothesis would not be rejected, even if there is some subset of series that are actually converging.

By its very nature the second group methods do not test for complete no convergence. When comparing their performance, one can notice that all these tests seem to be less powerful when the initial distance between series is high, when the speed of convergence is low or when less observations are available.

The data generation mechanism was particularly advantageous for unit root tests. As made clear from expression (3.17), a constant non-negative β results in a root that is smaller than 1.This is not a necessary condition for convergence, and more complex convergence processes could well mean that the power of DF type tests decrease. One has some ground to hope that the power of the Kalman filter method would be more robust to different data generation specifications.

Conclusion

Under a data generation mechanism that forces convergence of all series at the same speed, it comes without surprise that methods that use the whole cross-section are quite powerful in dismissing the no convergence null. The result of the simulations could also clarify some comparative advantages of methods that use two series only.

It would be interesting to study the robustness of all these methods if some departures from some simplifying assumptions of the data generation mechanism are considered. The next chapters cover the following situations:

- conditional convergence, meaning that the long run differences btween series are not necessarily zero (chapter 4);
- more than one attracting series (more than one convergence club, chapter 5);
- limited convergence, in the sense that only a subset of the series is converging (chapter 6);
- different speeds of convergence over time (chapter 7.)

The consideration of any of these departures could well mean that the power of some of the tests will decrease, unless they can be suitably adapted.

Appendix to chapter 3

Table 3.13
Real GDP per head in 1960
(in constant dollars, international prices, base 1985)

1 USA	9776	51 TAIWAN	1382
2 SWITZERLAND	9639	52 EL SALVADOR	1372
3 NEW ZEALAND	7920	53 SRI LANKA	1285
4 AUSTRALIA	7879	54 PARAGUAY	1215
5 SWEDEN	7492	55 DOMINICAN	1162
6 CANADA	7288	56 MADAGASCAR	1161
7 DENMARK	6751	57 JORDAN	1141
8 GERMANY, WEST	6637	58 MOZAMBIQUE	1129
9 UK	6548	59 PAPUA N.GUINEA	1128
10 VENEZUELA	6194	60 BENIN	1122
11 NETHERLANDS	6122	61 PHILIPPINES	1119
12 FRANCE	6013	62 BOLIVIA	1112
13 NORWAY	5685	63 TUNISIA	1088
14 BELGIUM	5583	64 CONGO	1059
15 TRINIDAD & TOBAGO	5577	65 SENEGAL	1016
16 FINLAND	5367	66 SOMALIA	1015
17 AUSTRIA	5176	67 HONDURAS	1007
18 ICELAND	5172	68 ZIMBABWE	990
19 ITALY	4636	69 IVORY COAST	975
20 URUGUAY	3829	70 ZAMBIA	944
21 ISRAEL	3322	71 THAILAND	929
22 ARGENTINA	3293	72 KOREA, REP.	907
23 SPAIN	3196	73 ANGOLA	879
24 IRELAND	3184	74 HAITI	873
25 PUERTO RICO	3069	75 GHANA	863
26 JAPAN	3033	76 MAURITANIA	862
27 CHILE	2893	77 BANGLADESH	798
28 MEXICO	2809	78 MOROCCO	790
29 IRAN	2535	79 EGYPT	770
30 CZECHOSLOVAKIA	2468	80 CAMEROON	701
31 HONG KONG	2210	81 UGANDA	680
32 SOUTH AFRICA	2109	82 CHAD	667
33 GREECE	2088	83 INDIA	665
34 COSTA RICA	2021	84 CENTRAL	661
35 YUGOSLAVIA	1955	85 KENYA	642
36 PERU	1917	86 INDONESIA	625
37 PORTUGAL	1869	87 PAKISTAN	618
38 JAMAICA	1788	88 BURUNDI	593
39 GABON	1786	89 NIGERIA	560
40 NAMIBIA	1772	90 BOTSWANA	552
41 BRAZIL	1758	91 RWANDA	519
42 ALGERIA	1717	92 NIGER	503
43 SINGAPORE	1712	93 MALI	499
44 COLOMBIA	1652	94 BURKINA FASO	473
45 GUATEMALA	1641	95 ZAIRE	459
46 TURKEY	1604	96 GUINEA	389
47 PANAMA	1520	97 MALAWI	371
48 SYRIA	1519	98 TOGO	364
49 ECUADOR	1433	99 MYANMAR	296
50 MALAYSIA	1397	100 LESOTHO	289

Source: Summers and Heston data set, mark 5.5, NBER gopher

Chapter 4
Conditional Convergence

Introduction

The previous chapter was based on a data generation process that implied unconditional convergence across all series. Several departures from this situation can be considered:

- only part of the series converge ("limited convergence");
- the series do not converge to the same attracting series ("convergence clubs");
- the series converge only conditionally, in the sense that their long run mean difference is constant but different from zero.

This chapter evaluates different methods of testing for convergence under a data generation process that induces conditional convergence across a hundred series. Therefore, it deals with the last departure considered above.

The notion of conditional convergence is explained in the first section. Next, the data generation process is described and some general properties of the generated data are assessed.

The following four sections deal with four different methods or families of methods of measuring convergence: initial value regressions, random field methods, Markov chains and time series methods. The chapter ends with a conclusion.

1.The conditional convergence idea

According to the earlier definition of convergence, two series are said to converge if their difference, in the long run, is stationary[1].

The difference mean needs not be equal to zero. If it is equal to zero, the series converge unconditionally. In the more general case, when the series differ by a constant (and a

[1]More precisely, their limit in probability is equal to a stationary series as time tends to infinity (see chapter 1 for more details.)

stationary error), the series converge conditionally. This is the case studied below.

The term "conditional convergence" arises in the convergence theoretical and empirical literature and is often associated with some version of the neoclassical growth model. In this family of models, every country product per head grows at the same rate in the long run (the exogenous rate of technological progress.) Nevertheless, this does not imply that product levels are the same for each country. In the long run, countries may be positioned in different steady-states, their long run differences being constant. Empirically, these different steady-states are usually proxied by a set of conditioning variables[2].

As explained in Chapter 2, other income convergence models display the conditional convergence property. Moreover, the conditional convergence concept can be applied to other series where there is a convergence issue, like prices, inflation or interest rates.

In the next section, the conditional convergence model that is the base for the experiments is presented and some basic properties of the generated data are discussed.

2. Simulation of a conditional convergence process

2.1 The data generation process

This model is an extension of the one used before, to study unconditional convergence. A hundred "GDP per head-like" series are generated for a hundred countries[3]. The initial values are given by the natural logarithms of GDPs per head in 1960. The US series is the attracting series and is generated according to the following equation:

$$x_{1,t}=x_{1,t-1}+g+\eta_{1,t}+\Delta\epsilon_{1,t}, \tag{4.1}$$

where g is the US average growth rate between 1961 and 1989:

[2]For a derivation of this result in neoclassical growth models with optimising consumer and human capital see chapter 1. For a survey of some empirical results, see chapter 2.

[3]See the chapter on unconditional convergence (chapter 3) for the list of 100 countries and respective GDP's per head in 1960.

$$g = 0.0217. \tag{4.2}$$

η is a permanent shock and ϵ a transitory one. Both are mean zero normally distributed uncorrelated random variables. Its variance is set according to the verified variance of the US growth rate:

$$Var(\epsilon) = Var(\eta) = \frac{1}{3} Var(US\ growth\ rate), \tag{4.3}$$

with:

$$Var(USgrowthrate) = 0.00070107. \tag{4.4}$$

The 99 attracted variables were generated obeying the following:

$$x_{i,t} = x_{i,t-1} + g_i + \beta(x_{1,t-1} - x_{i,t-1}) + \eta_{i,t} + \Delta\epsilon_{i,t}, \qquad i = 2, 3, \ldots, 100. \tag{4.5}$$

β (the speed of convergence) is non-negative and smaller than 1. Again, η and ϵ are mean zero normally distributed uncorrelated random variables.

In the long run, each attracted series grows at the rate g, and the variable g_i only determines the steady-state for each series. It can be shown that the long run difference between the attracting series and series i (d_i) is equal to:

$$\overline{d}_i = \frac{g - g_i}{\beta}, \tag{4.6}$$

so that the two series converge unconditionally if g_i is equal to g and conditionally otherwise. If g_i is higher than g, series i tends to be positioned above the attracting series, in the long run. Conversely, a g_i that is smaller than g implies a steady-state for series i that is lower than the attracting series steady-state.

Note that β is fixed from the beginning and equal for every series. One only has to determine the g_is to be able to simulate the set of a hundred series.

It was arbitrarily assumed that each series steady-state difference was equal to the first period steady-state.

2.2 Some properties of generated data

Artificial data generated along the lines described above displayed some properties that were similar to real data from the Summers-Heston database. Namely:

i) The cross-section dispersion of incomes per head increases over time, as it becomes clear from Figure 4.1.

Artificial data corresponds to a value of β equal to 0.05 and the cross-section dispersion is an average after 1000 replications.

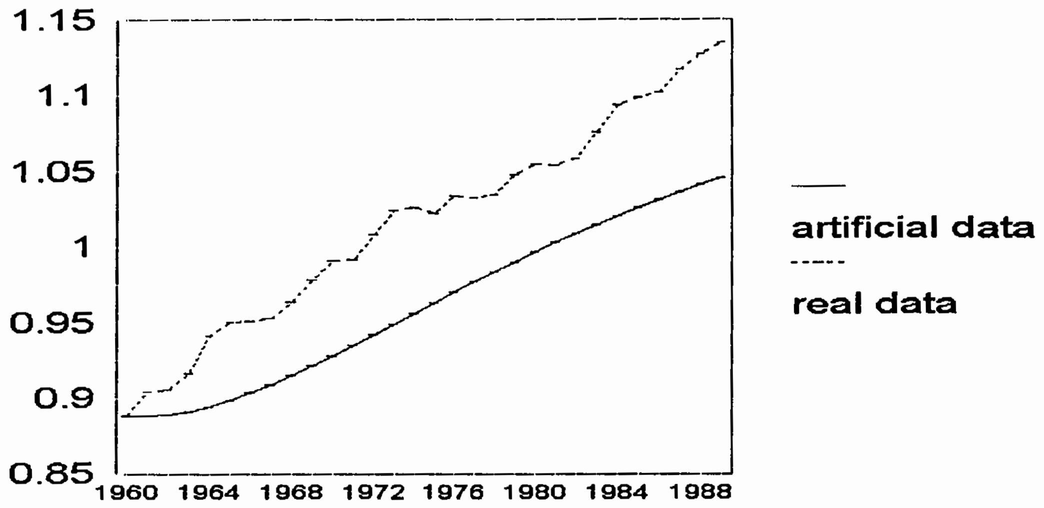

Figure 4.1
Standard deviation for artificial and real data

Note that this happens even if all the countries are conditionally converging to the leader (in the " β convergence sense") . In " σ -convergence " terms, it can be said that the initial cross-section dispersion is smaller than its steady-state level.

ii) Some countries seem to be catching up, while others follow behind.

Table 4.1 compares the maximum and minimum growth rate in real and artificial data. In both cases some countries grow at a much faster rate than the leader while others fall behind and have negative growth[4].

Table 4.1
Growth rates in real and artificial data compared

	real data (Summers and Heston)	artificial data
maximum average annual growth rate	6.63%	7.87%
minimum average annual growth rate	-1.94%	-2.57%

3. Initial value regressions

To assess the power and properties of initial value regressions, different values for the parameters and specifications of the regressions were experimented. In every case, a thousand replications of the hundred conditionally converging series were simulated.

The general form of the initial value regression is:

$$\frac{x_{it}-x_{i1}}{T-1} = b_0 + b_1.c_i + b_2.x_{i0} + error, \qquad i=1,2,\ldots,100, \qquad \textbf{(4.10)}$$

where c_i is a conditioning variable.

Two different number of periods T were tried: a "long" one (100 periods) and a "short" one (30 periods.) The short time span is close to the one used in empirical conditional convergence regression studies. Also, three different "conditioning" variables c were generated and included in the regression: the first is perfectly correlated to the steady-state variable $\overline{d}_i$, the second displays a correlation coefficient to the steady-state equal to 0.8, and for the third one the correlation is equal to 0.6.

[4]Bear in mind that the leader is growing at a rate equal to 2.17 % per year.

Table 4.2 displays some results for the case where the correlation between the conditioning variable and the steady-state variable (ρ) is equal to one.

Table 4.2
Initial value regressions
T=100, ρ=1.0
(1000 replications)

β (speed of convergence)	% of no convergence rejections 5 % (1%)	% of no conditional convergence rejections 5 % (1%)	Max $\hat{\beta}$	Min $\hat{\beta}$	Average $\hat{\beta}$	% of successes
0.00	5.4 (1.3)	4.4 (0.7)	0.0008	-0.0008	0.000	100.0
0.02	100.0 (100.0)	100.0 (100.0)	0.0236	0.0174	0.0200	100.0
0.10	100.0 (100.0)	100.0 (100.0)	0.10072	0.0388	0.0556	51.6

The null hypothesis of no convergence is rejected when the t-value of the initial value coefficient is statistically not different from zero. Under the "% of no conditional convergence rejections" the percentage of times the coefficient on c_i was significant is presented. An estimate of β ($\hat{\beta}$) is obtained from the estimated coefficient on the initial value:

$$\hat{\beta} = 1 - [(T-1).\hat{b}_2 + 1]^{\frac{1}{t-1}}. \tag{4.11}$$

It is possible that the expression $[(T-1).\hat{b}_2+1]$ turns out to be negative. In that case, it is not possible to provide an estimate of β. When this estimate exists, it is recorded as a "success" (see last column it Table 4.2.)

As expected, when β=0.0 , i.e., when there is no convergence, the no convergence hypothesis is rejected at percentages that are close to the significance level. Also, since the growth rates g_i were made equal to g, it is not surprising that most of the times the coefficient on the conditioning variable is not significant.

When there is conditional convergence (last two lines), the no convergence hypothesis was always rejected. Moreover, when the speed of convergence is low (equal to 0.02), it was possible to estimate it quite accurately: $\hat{\beta}$ was always between 0.0174 and 0.0236, and this estimator proved to be a centred one. The same can not be said when the speed is higher (equal to 0.10.) In this last case, $\hat{\beta}$ was biased towards zero, and in half of the times it was not possible to calculate it.

In every case, the "no conditional convergence hypothesis" was rejected, meaning that the conditioning variable coefficient was always found to be significant.

This last result comes unsurprisingly, since the correlation between the conditioning and the steady-state variables is perfect. In practice, this is not going to happen, and the researcher has to rely on proxies for different steady-states. Having this in mind, two other sets of replications were programmed, with conditioning variables as regressors that were only partially correlated to the steady-state. Summary results are presented in Table 4.3.

Table 4.3
Initial value regressions
T=100, β=0.02
(1000 replications)

ρ (correlation between conditioning variable and steady-state)	**% of no convergence rejections 5 % (1%)**	**% of no conditional convergence rejections 5 % (1%)**	**Max** $\hat{\beta}$	**Min** $\hat{\beta}$	**Average** $\hat{\beta}$	**% of successes**
0.8	100.0 (100.0)	100.0 (100.0)	0.0123	0.00216	0.00739	100.0
0.6	53.3 (27.2)	97.5 (90.5)	0.00434	-0.0033	0.00147	100.0

Two interesting remarks can be made about these results. Firstly, the speed of convergence was always underestimated. Secondly, the no convergence hypothesis was not rejected more than a half of the times, when the correlation was equal to 0.6. In other words, the power of the test declines dramatically if one does not have a good proxy for the steady-state levels[5].

When the speed of convergence is equal to 0.05 and T=30 (Table 4.4), the estimated speed of convergence is equal to 0.0215 only, on average, if the correlation between conditioning and steady-state variables is equal to 0.8. This average estimate is very close to the empirical results by Barro and Sala-i-Martin (1992a, 1992b, 1995). Note that this is the case that corresponds to Figure 4.1 and to Table 4.1.

Table 4.4
Initial value regressions
T=30, β=0.05
(1000 replications)

ρ	% of no convergence rejections 5 % (1%)	% of no conditional convergence rejections 5 % (1%)	Max $\hat{\beta}$	Min $\hat{\beta}$	Average $\hat{\beta}$	% of successes
0.8	100.0 (100.0)	100.0 (100.0)	0.0368	0.0105	0.0215	100.0

4. Random field methods

4.1 The random field regressions

This method uses all the data in the sample and is basically a time series-cross-section regression. The coefficient b in the following equation is estimated by OLS:

$$d_{ij} = b.d_{i,j-1} + error, \quad i=2,...,N; \quad j=1,..,T, \tag{4.12}$$

N being the total number of series (a hundred in this case) and T the total number of periods. The no convergence hypothesis is associated with a true value of 1 for b. This means that d_{ij} has a unit root. This is not a pure univariate time series process, and the OLS estimate of b

[5]Results for a smaller number of periods (T=30) were qualitatively very similar and are not presented here.

Table 4.6
Random field regressions
T=100
(1000 replications)

β	% of no convergence rejections 5 % (1%)	Max $\hat{b}$	Min $\hat{b}$	Average $\hat{b}$
0.02	0.0 (0.0)	1.0016	0.999	1.000
0.10	0.0 (0.0)	1.0016	0.999	1.000

4.3 Why is the power of random field tests so low?

Random field regressions are not suited to detect a *conditional* convergence situation. They can dismiss the no convergence hypothesis when the true hypothesis is that all the series are converging *unconditionally*, this is to say, when all the series converge to the same level in the long run.[6]

To understand the reason why this is so, one has to take the difference between x_{1t} and x_{it} from equations (4.1) and (4.5):

$$d_{it} = (1-\beta).d_{i,t-1} + g - g_i + error, \tag{4.14}$$

and substitute g_i from (4.6):

$$d_{it} = \beta.\overline{d}_i + (1-\beta).d_{i,t-1} + error. \tag{4.15}$$

In equation (4.15) there is a country-specific effect that is not taken into account in the

[6]This was shown in the previous chapter.

random-fields regression. When this country-specific effect did not exist (when $\overline{d_i}=0$, for every i), the random field estimator b_{ols} proved to be a good estimator for $1-\beta$. Note that the introduction of a constant in the regression does not solve the problem. One possible solution would be to resort to panel data techniques and include a country-specific dummy.

5. Markov chains

The Markov chains method was proposed by Quah (1993a) and described in the last chapter.

The results do not differ much when different speeds of convergence or number of periods are considered. Therefore, results of one simulation where β=0.05 and T=100 are presented.

Firstly, the one-period transition matrix is estimated (M_1.) Remember that element m_{ij} is the probability of a series in the j-th quantile in one period to swap into the i-th quantile in the following period.

$$M_1 = \begin{pmatrix} 0.979 & 0.0346 & 0.0 & 0.0 & 0.0 \\ 0.0207 & 0.942 & 0.0170 & 0.0 & 0.0 \\ 0.0 & 0.0235 & 0.956 & 0.0196 & 0.0 \\ 0.0 & 0.0 & 0.0269 & 0.969 & 0.0189 \\ 0.0 & 0.0 & 0.0 & 0.0111 & 0.981 \end{pmatrix} \qquad \textbf{(4.16)}$$

The probability of a country swapping into a neighbouring quantile, although positive, is small. Namely, it is smaller than the same probability in a context of unconditional convergence (compare to the same matrix in the chapter on unconditional convergence.) The diagonal elements, which denote the probability of a country to remain in the same quantile from one period to the other, are always greater then 0.94

The long run distribution implied by matrix M_1 is expressed by vector F_∞:

$$F_\infty = \begin{pmatrix} 0.237 \\ 0.142 \\ 0.196 \\ 0.268 \\ 0.157 \end{pmatrix} . \tag{4.17}$$

In the long run, countries are spread along the quantiles. This is not the pattern associated to unconditional convergence. Under unconditional convergence, the series would be concentrated in the middle quantiles.

Identical conclusion result from the 100-period transition matrix and corresponding long run distribution. These are reproduced below:

$$M_{100} = \begin{pmatrix} 0.667 & 0.441 & 0.0833 & 0.0 & 0.0 \\ 0.167 & 0.206 & 0.292 & 0.0 & 0.0 \\ 0.167 & 0.235 & 0.292 & 0.333 & 0.0556 \\ 0.0 & 0.0882 & 0.292 & 0.583 & 0.222 \\ 0.0 & 0.0294 & 0.0417 & 0.0833 & 0.722 \end{pmatrix} \tag{4.18}$$

$$F_\infty = \begin{pmatrix} 0.239 \\ 0.136 \\ 0.235 \\ 0.262 \\ 0.128 \end{pmatrix} . \tag{4.19}$$

As it was expected, the diagonal elements in the transition matrix are now smaller: in a hundred periods time, the probability of swapping into another quantile is higher. The long run distribution shares the same properties with the one derived from the one-period matrix: the incomes distribution is bimodal with a concentration of countries in the lowest quantiles.

In both cases, a conditional convergence simulation induced results that were considerably different from the unconditional convergence ones. Moreover, the above results were similar to the ones Quah reports using the Summers-Heston database.

6. Time series methods

This category includes methods that test for convergence using only two series at a time. Such methods were already presented in the previous chapter. Briefly, they are:

i) The Dickey-Fuller method. The difference between two series is tested for stationarity using the Augmented Dickey-Fuller statistic.

ii) The Kalman filter method. The specification for the measurement equation is:

$$d_{it}=\alpha_{it}+\epsilon_{it}, \qquad \textbf{(4.20)}$$

the transition equation being:

$$\alpha_{it}=\alpha_{i,t-1}+\mu_{it}, \qquad \textbf{(4.21)}$$

where $Var(\mu_{it}) = \phi_i^2 Var(\mu_{i,t-1})$. If ϕ_i is estimated to be significantly smaller than 1, the no convergence hypothesis is rejected.

In both cases, it can be shown that estimation results depend only on the initial distance from the steady-state. This means that if the difference from the attractor of series i is equal to 3 in time 1, the steady-state difference being equal to 2, and if the values for series j are 4 and 3, respectively, than the tests for convergence between series i and j and the attracting series will have the same power.

Results from the previous chapter may therefore be applied here. The no convergence hypothesis will be rejected when the speed of convergence and the distance from the steady-state are higher.

Conclusion

All the simulations in chapter imposed conditional convergence of every series, at the same speed, from the outset. Two methods that performed well under an unconditional convergence framework (random field regressions and Markov chains) lose their power dramatically when in presence of a conditional convergence situation.

The initial value regressions proved to very successful in an idealized world where not only every series converge at the same speed, but also a perfectly correlated variable to each steady-state level is known by the researcher. When this last condition fails, and the steady-state level is only proxied with an error, the no convergence hypothesis is often not rejected and the speed of convergence tends to be underestimated.

It was possible to generate data with all the series conditionally converging at the same rate to the same leading series that displayed some resemblance to real data :

- the cross-section dispersion of GDP's per head increases in a thirty-year period;

- some countries grow at a considerable higher rate than the leader while others exhibit negative growth;

- initial value regressions including conditioning variables that are correlated to the steady-state level rightly dismiss the no convergence hypothesis;

- random field regressions fail to recognise the convergence process going on;

- Markov chain methods allow the researcher to conclude that the world is becoming more divided between richer and poorer countries.

Time series methods all rely on the behaviour of the differences between any two series and their power depends on the initial distance from the steady-state. In a context where all the series are conditionally converging, these methods are more likely not to reject the null of no convergence when the series start close to the steady-state.

Chapter 5
Convergence Clubs

Introduction

In chapter 3 different methods to measure convergence were presented and its performance evaluated when the true data generation process implied unconditional convergence across all series. In chapter 4 one departure from this convergence framework was considered: the series were supposed to converge conditionally, so that their long run difference was other than zero. In both chapters, series were supposed to converge to the same leading series. The "transition property" applies to convergence: if series A converges to series B, and if series B converges to series C, then series A converges to series C. Therefore, all the series converge to each other in the data generation processes considered until now.

The data generation process studied in this chapter implies that it is not any longer true that all the series converge to each other. By introducing different leading series, different convergence clubs are contemplated. Each series belongs to one and only one club, according to the leading series it converges to. Only two clubs are considered and series converge unconditionally to their leader. More complex situations would probably obscure the main points being made without any particular gain.

The first section in this chapter describes the data generation process and, particularly, the different ways of generating clubs. In the simulations presented, belonging to a club is either perfectly correlated to the initial level of income, or only partially related to this variable, or completely independent from it.

The misleading information that can arise from the application of unmodified initial value ("Barro-type") or random field regressions is illustrated in section 2.

Alternative existing methods that take into account the possible existence of more than one club also have their failures. Quandt-like tests, proposed by Durlauf and Johnson (1992) have some troublesome results when the composition of the clubs is not completely correlated to initial income. This is shown in section 3. That the same can be said of Markov chains

- (i) The first half of the sample (the 50 richest countries) converge to country 1 and the other half converge to country 51. This pattern is labelled as "highly related to income."

- (ii) There is no relation between initial income and income group. "Odd number" countries converge to country 1 and "even number" countries converge to country 51. This situation is labelled as "unrelated to income."

-(iii) There is only a partial relation between initial income and convergence to the richest club. A country is more likely to belong to the first club if its initial income is high, but some countries that are rich in the beginning fall later on into the "second division." This is a "partially related to income" pattern and the composition of each club is as follows:

Club 1: countries 1 to 8, 10 to 19, 21, 23 to 31, 33, 35, 37, 39, 40, 43, 44, 47, 50, 54, 58, 61, 63, 66, 68, 72, 73, 77, 79, 83, 88, 92, 93 and 100.

Club 2: all the remaining series.

2. Misleading information from Barro-type and random field regressions

Durlauf and Johnson (1992) observed that initial value (or "Barro-type") regressions could lead to misleading information in what concerns convergence. They argue that the existence of convergence clubs can lead to estimates of the parameter on the initial value for income that are negative. That this is the case is shown next.

A period of thirty years (roughly the period that is considered in studies that use the Summers-Heston data base) and the different patterns for the location of clubs that were described in the last section were considered.

The initial value unconditional regressions results are summarised in the Table 5.1.

Table 5.1
Initial value regressions
2 clubs
β=0.02
(1000 replications)

position of clubs:	% of no convergence rejections 5 % (1%)	average estimated β
highly related to income	0.1 (0.0)	0.000
unrelated to income	100.0 (100.0)	0.0199
partially related to income	100.0 (100.0)	0.007

The estimated regressions are of the type:

$$\frac{x_{i,T}-x_{i,1}}{T-1} = constant + \frac{(1-\beta)^{T-1}-1}{T-1}x_{i,1}+v_{i,T}. \qquad \textbf{(5.2)}$$

Equation (5.2) is estimated using OLS. The estimate for β is computed from the estimated coefficient on the initial value. "No convergence" is rejected when this last coefficient is statistically negative.

Although initial value regression fail to detect any kind of convergence in the highly related to income situation, it is interesting to note that in the other two cases the no convergence hypothesis was always rejected. In the "unrelated to income" pattern, the average estimated β was very close to its true value in the simulations[2] (0.02.) These results could mislead the researcher into accepting a hypothesis of full convergence, the reality being quite different.

[2]This happens because the omitted variable (a binary dummy that varies according to the club each series belongs to) is orthogonal to the regressor (initial income.) See section 3, below.

Something similar is apparent if random field regressions are appraised. Consider Table 5.2, where results obtained using this last method are summarised:

Table 5.2
Random field regressions
2 clubs
β=0.02
(1000 replications)

position of clubs:	% of no convergence rejections 5 % (1%)	average estimated b
highly related to income	3.8 (0.0)	0.995
unrelated to income	47.7 (1.2)	0.992
partially related to income	16.4 (0.1)	0.994

The no convergence hypothesis is not rejected so often, but it is still possible to reject it. The nature of the convergence process going on is not apparent from the results, though.

3. Quandt tests

Durlauf and Johnson (1992) propose several alternatives to test for the existence of convergence clubs[3]. They are all based on the same idea: the existence of clubs implies heterogeneity in initial value regression parameters. This can be illustrated with the data generation process. After some algebra, the following expression is derived from (5.1):

$$\frac{x_{i,T}-x_{i,1}}{T-1} = \frac{y_{j,T}-(1-\beta)^{T-1}y_{j,1}}{T-1} + \frac{(1-\beta)^{T-1}-1}{T-1}x_{i,1} + (T-1)^{-1}\sum_{t=2}^{T}(1-\beta)^{T-t}u_{i,t}, \quad \textbf{(5.3)}$$

with

[3]Their work was reviewed in chapter 2

$$u_{i,t} = \eta_{y_j,t} - \eta_{i,t} + \Delta\epsilon_{y_j,t} - \Delta\epsilon_{i,t}, \quad \textbf{(5.4)}$$

and where j equals 1 or 2, if X_i converges to Y_1 or to Y_2, respectively. Equation (5.3) is an initial value cross-section equation. The intercept is different according to the group. If this equation is estimated without considering the heterogeneity in the intercept, as it was done in the last section, it is very likely that the estimates are inconsistent. If the researcher knew the composition of each group beforehand, the coefficient of the initial value could be consistently estimated by introducing a binary dummy variable in the initial value estimated equation.

The next task is to identify the members of each club. Durlauf and Johnson propose two methods that allow for the endogenous determination of clubs - the single control variable and the regression tree methods.

The single control variable method is examined here. It is an application of Quandt's (1958) technique, generally used to test for a structural break in a time series regression. The control variable and the number of splits are exogenously chosen. If n_s is the number of splits and c is the control variable, then n_s+1 clubs are considered and one separate regression is estimated for each of them. Club 1 includes countries for which c is equal or less then c_1, club n_s+1 comprises countries for which c is higher than c_{ns+1}, and club i, with $1 < i < n_s+1$, includes countries for which c falls between c_i (*exclusive*) and c_{i+1} (*inclusive*.) The location of the splits is endogenously determined by choosing the one that maximises the likelihood function.

In the simulations, the number of splits equals 1 (this means that the number of clubs was made equal to its "true" value.) The control variable is income per capita. Remembering that the number of countries is equal to 100, the log-likelihood function becomes:

$$\log L = C - s.\log\hat{\sigma}_1 - (100-s).\log\hat{\sigma}_2 \quad \textbf{(5.5)}$$

where s is the number of countries in club 1 (100-s is the number of countries in club 2), $\hat{\sigma}_i$ is the estimated standard deviation of the error term in the ordinary least squares regression for club i and C is a constant. The optimal value for s maximises log L (feasibility of ordinary least squares have to be assured for each sample.)

Although Durlauf and Johnson do not go this far in their paper[4], a complete test involves a proper statistic and corresponding critical values so that the hypothesis of "no split" is tested.

Let log L_r be the restricted log-likelihood, i. e. the log-likelihood function that corresponds to the hypothesis that all the countries belong to the same club. L_r is given by:

$$\log L_r = C - 100.\log\hat{\sigma}, \tag{5.6}$$

where $\hat{\sigma}$ is the estimated standard deviation from the restricted OLS regression (i. e. considering the whole sample) and C is the same constant as in (5.5).

A log-likelihood ratio statistic is usually given by:

$$LR = 2.(\log(\textit{unrestricted likelihood}) - \log(\textit{restricted likelihood})). \tag{5.7}$$

In this case the LR statistic becomes Quandt (1958) statistic and is equal to:

$$Q_r = 2.(100.\log\hat{\sigma} - s.\log\hat{\sigma}_1 - (100-s)\log\hat{\sigma}_2). \tag{5.8}$$

Q_r does not follow the usual χ^2 distribution and is poorly approximated by it, as Quandt (1958, 1960) notes. Instead, it depends on the sample size as well as on the number of regressors.

Therefore, critical values under the null of no different clubs (homogeneity in the coefficients) were computed using Monte Carlo methods[5]. These critical values are given in Table 5.3.

---

[4]These authors only estimate the position of the splits but do not compute their statistical significance.

[5]Deutsch (1992) provides tables for the cumulative empirical distribution of the statistic. They are close to the

Table 5.3
Quandt statistic
Critical values
(1000 replications)

significance level	critical values
0.90	11.0464
0.95	12.8729
0.99	16.6630

The actual results when the data generating process implied the existence of two clubs are summarised in Table 5.4.

Table 5.4
Quandt test
2 clubs
β=0.02
(1000 replications)

position of clubs	% of "no clubs" rejections 5 % (1 %)	mode of c_1 (split position) (number of occurrences)
highly related to income	100.0 (100.0)	50 (1000)
unrelated to income	0.0 (0.0)	59 (153)
partially related to income	99.9 (96.2)	8 (794)

When the clubs' position is "highly related to income" (the first club comprises the 50 richest countries in the first period), there is no problem at all in rejecting the null hypothesis of "no clubs" and in identifying the clubs. In fact, the null hypothesis was always rejected. Moreover, the split was always (a thousand times) located in series 50, so that club membership was always correct.

Completely opposite results arise when club membership is "unrelated to income." In this case, no heterogeneity is detected by this method and the null of no clubs is never rejected.

When club membership is partially related to income, the "no clubs" null is almost always rejected. It is interesting to note that in 79.4 percent of the times this method includes series 1 to 8 in the first club only, even if this club includes 52 countries in the simulated reality (37 of them being located in the first half of the sample.) The fact that series 9 belongs to the second club is responsible for this result, even if series 1 to 8 and 10 to 19 converge to the first series. This method seems to provide poor information in the probably more realistic situation of a control variable that is imperfectly correlated to club membership.

4. Markov chains

This method, proposed by Quah (1993a), was already described in chapter 3 and was also implemented in chapter 4. In all the simulations considered here the total number of periods equals 30 and the speed of convergence is equal to 2 percent.

Recall that the method implies the estimation of a transition matrix. The one-period transition matrix (M_1) is derived from the period to period observed transitions from one quantile to the other[6]. Element m_{ij} is the probability of a series in the j-th quantile in one period to swap into the i-th quantile in the following period.

From each transition matrix it is possible to compute the log-run implied distribution of GDPs per head across countries. This distribution is summarised in vector F_∞ ; element f_i is the percentage of countries in the corresponding quantile. The lowest quantiles correspond to poorer countries.

Here, Markov chain methods are applied to the three different settings that include the existence of two clubs.

[6]The 30-period transition matrix (M_{30}) is similar in construction and meaning, but computed from the full-period observed transitions. It is not presented here because results were qualitatively similar.

4.1 Position of clubs highly related to income

In a case where the first division comprises the 50 countries that were richer in the beginning[7], the one-period transition matrix is the following:

$$M_1 = \begin{pmatrix} 0.975 & 0.0193 & 0.0 & 0.0 & 0.0 \\ 0.0254 & 0.980 & 0.00803 & 0.0 & 0.0 \\ 0.0 & 0.000985 & 0.888 & 0.0105 & 0.0 \\ 0.0 & 0.0 & 0.104 & 0.990 & 0.00192 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.998 \end{pmatrix} \qquad \textbf{(5.9)}$$

The most striking figure of this matrix is probably the low value for the central diagonal element, when compared to the other diagonal entries. This means that only 88 percent of the "middle class" countries are expected to remain in that group during the next period. More than 10 percent change into the richer adjacent quantile. The other diagonal elements are much closer to 1, so that countries in other quantiles tend not to change their relative position.

In the long term, this tendency for the middle group to vanish is clearly expressed in matrix F_∞ :

$$F_\infty = \begin{pmatrix} 0.245 \\ 0.322 \\ 0.0395 \\ 0.393 \\ 0.0 \end{pmatrix} . \qquad \textbf{(5.10)}$$

The long run distribution is bimodal, with a concentration of countries in the second and fourth quantiles. Although these results depend clearly on the initial definition for the quantiles, they correspond to what one would expect from the imposed pattern of convergence. The long run distribution expresses a divided world with a concentration of countries both in the poorest and the richest halves.

---

[7]This is the result of one replication. Other replications displayed very similar results.

4.2 Position of clubs completely unrelated to income

In this case, the one-period transition matrix is qualitatively different from the previous one:

$$M_1 = \begin{pmatrix} 0.910 & 0.0124 & 0.0 & 0.0 & 0.0 \\ 0.0902 & 0.956 & 0.0127 & 0.0 & 0.0 \\ 0.0 & 0.0317 & 0.961 & 0.0274 & 0.0 \\ 0.0 & 0.0 & 0.0265 & 0.970 & 0.00213 \\ 0.0 & 0.0 & 0.0 & 0.00230 & 0.979 \end{pmatrix}. \tag{5.11}$$

Take, for instance, the third quantile; it is comprised by countries that have a GDP per head that is lesser than the average but higher than half of the average. These countries have a high probability of staying in the same quantile the following period (96.1 per cent), and the chance of swapping into to the fourth quantile (which comprises richer countries) is greater than the possibility of falling into the company of poorer countries. The same could be said of the second quantile, and the inverse applies to the fourth one. It is worth noting that the highest probability of not staying in the same quantile is to be found among the poorest countries. This tendency for countries to be attracted into the middle-income quantiles is reflected in the implied long run distribution expressed by the following vector:

$$F_\infty = \begin{pmatrix} 0.0219 \\ 0.159 \\ 0.396 \\ 0.382 \\ 0.0412 \end{pmatrix}. \tag{5.12}$$

This long run distribution would be generally interpreted as a tendency for countries to cluster around the median quantiles, the relative number of very rich and very poor countries being very small (less than 7 percent of countries fall into the extreme quantiles.) This is at odds with what the data generation process implies. This data was generated by a process that implied two different clubs, and their relative levels of income are not at all different from the situation described in a.) The only difference is the starting point. Here, some poor countries join the richer club and there are countries that start rich but get

Conclusion

Existing cross-section tests for convergence are not very successful in dealing with convergence clubs.

Initial value and random field regressions are limited from their inception. When there are two or more convergence clubs, results from performing a regression using all the series have to be read carefully. If the no convergence hypothesis is rejected, the researcher should take care in not concluding immediately that all the series are converging to the same leading series. Moreover, if "no convergence" is not rejected, this could be due to the fact that the series are converging to different attractors and that there is more than one club.

Quandt tests are a trial to make the choice of clubs endogenous. The number of clubs is defined *a priori*. Even if this choice is adequate (equal to the true number of clubs), it is possible that the composition of the clubs is very different from the true one. This will happen if the correlation between the control variable and the belonging to a specific club is not very high.

Something similar happens when applying the Markov chains estimation method; when the composition of clubs is not completely related to income, this method fails to detect the underlying bimodal distribution in the true data.

Time series were not examined in this chapter. Note that these methods use two series at a time. Therefore, belonging of a specific series X to a club A can be directly tested in the following way: choose a series Y from A and test for convergence between series X and Y. If X converges to Y, than X can be included in club A.

Another possible procedure would be to follow Hall, Robertson and Wickens (1993). Considering two series Y and Z that belong to two different clubs, convergence of X to Y could be tested by estimating the following equation using the Kalman filter[8]:

---

[8]See chapter 2 for more details on this model.

$$Y_t - X_t = \alpha_t + \beta_t . (Y_t - Z_t). \quad \textbf{(5.15)}$$

If β and the variances of α and β tend to zero, there is evidence of convergence between X and Y. If both variances tend to zero and β tends to 1, X converges to Z.

Chapter 6
Limited Convergence

Introduction

The underlying DGP in chapter 5 implied the existence of two different clubs. In this chapter another departure from a situation where all the series converge to the same steady-state is considered. Only a part of the series converge unconditionally to one leading series. In this respect, the situation within this converging group is not very different from the one analysed in chapter 3. The remaining series do not converge at all. Instead, they grow at an average rate that is equal to the leader's growth rate. Since each series is affected by unrelated permanent shocks, the differences between non converging series are random walks.

The first section describes how this limited convergence world is generated. Initial value and random field regressions fail in the presence of a limited convergence framework. This point is made in section 2. Difficulties in assessing limited convergence and in identifying the converging set of countries when using Markov chains are dealt with in section 3. As usual, the chapter ends with a conclusion.

1. Generating limited convergence

In the limited convergence data generation process only a proportion p_c of the series converge unconditionally. The remaining series do not converge to any particular steady-state.

The data generation process (DGP) for unconditionally converging GDPs per head was presented in chapter 3. This DGP is used here to generate the subset of converging series. The remaining series are created from an equation similar to the one used for the attracting series:

$$x_{j,t}=x_{j,t-1}+g+\eta_{j,t}+\Delta\epsilon_{j,t}, \qquad \textbf{(6.1)}$$

where j is an index defining a non converging series.

A hundred series are generated in which series 1 is the attracting series. Series are ordered according to the initial value, so that series with a lower index start from a higher point. The subset of converging series can either be:

- located in the beginning of the sample, so that the converging series are the richest ones in the beginning of the period;

- uniformly distributed along the sample. If a quarter of the series is to converge, than the convergence subset comprises series 2, 6, 10, etc.

- positioned at the end of the sample. The poorest countries in the world converge to the richest one, and the ones in the middle are left wandering around[1].

2. Initial value and random field methods

2.1 These methods are condemned to fail: theory

Some authors have already noted that the existence of "convergence clubs" could be responsible for a "conditional convergence" result when running an initial value regression that includes some conditioning variables[2]. The basic idea is simple and applicable to a limited unconditional convergence framework; if a subset of series is converging to an attractor, there is a positive correlation between initial values and growth rates, even if the remaining series evolution is completely independent from the attractor. This positive correlation will possibly show in a negative (and statistically significant) coefficient when growth rates are cross-sectionally regressed on initial values.

More generally, one can say that initial value regressions are inappropriate to deal with a limited convergence situation because the information they provide is usually not enough to allow the researcher to formulate a meaningful conclusion. If a negative coefficient on initial income is found, one does not know whether *all* the series are converging, or whether only

[1]The sceptic reader is probably thinking this is an odd hypothesis to work with. However, it can be argued that:

a) convergence can be conditional, so that it is possible that the converging countries are not closer together in the long run;

b) this hypothesis is useful to understand the properties of some methods to measure convergence.

[2]See Durlauf and Johnson (1992) and Quah (1994).

a part of them is. If no statistically significant relationship is uncovered, it is still possible that there is a subset of series that are converging, the size of the subset being too small or the convergence speed being too slow to be revealed in a statistically negative coefficient.

The last criticism also pertains to panel data or random field regressions. These methods impose restrictions across series that are not verified when limited convergence is at work. In chapter 4 Monte Carlo evidence is presented that shows that random field tests fail to reject the null of no convergence when *all* the series converge conditionally. Then, the estimated equation imposed homogeneity in the intercept when the data generating process implied different intercepts across series. Here, heterogeneity is extended to the slopes.

Record that the random field estimated equation is:

$$d_{it} = b.d_{i,t-1} + error, \qquad \textbf{(6.2)}$$

where d_{it} denotes the difference between series i and the attractor. Under limited unconditional convergence, the coefficient b is different for the two sets of series. If series i belongs to the subset of converging series, its b will be smaller that one, so that the difference is stationary. If the series is part of the no convergence group, its b is equal to one, the difference being a random walk. Again, under limited convergence, it is both possible to have an estimated b that is smaller or statistically equal to one, depending on the size of the convergence group and on the speed of convergence.

2.2 These methods are condemned to fail: Monte Carlo evidence

Artificial data were generated where the proportion of countries that converge unconditionally was first equal to 0.3 and later equal to 0.6.

Results were not very disparate when different locations for the converging club were considered. Results for initial value regressions concerning the case where the converging series are placed at the beginning of the sample are reproduced in Table 6.1.

Table 6.1
Initial value regressions
T=30, β=0.06
Converging series uniformly distributed across the sample
(1000 replications)

proportion of converging countries:	Average $\hat{\beta}$	% of no convergence rejections (5 % level)
p=0.3	0.00921	98.6
p=0.6	0.0176	100.0
p=1.0	0.0600	100.0

One interesting finding is that no convergence was almost always rejected, even when converging series are less than a third of the total. The true convergence speed was made equal to 0.06. As one could expect, the estimated speed of convergence was always between 0 and 0.06. It is greater when the proportion of converging countries is higher and smaller when the correlation of conditioning variables to the steady-state is lower. Its average was equal to its true value when all series were converging.

Table 6.2
Random field regressions
T=30, β=0.06
(1000 replications)

	Average $\hat{b}$	% of no convergence rejections (5 % level)
p=0.3	0.999	0.1
p=0.6	0.993	28.4
p=1.0	0.960	100.0

These results suggest that findings reported by Barro and Sala-i-Martin (1992a, 1992b, 1995) and by Mankiw, Romer and Weil (1992) can arise from a situation of limited convergence instead of a situation where *all* the countries are conditionally converging.

Results for random field regressions are summarised in Table 6.2.

No convergence is almost never rejected when the convergence club is smaller. When the convergence club gets larger, the no convergence hypothesis is more likely to be rejected.

Recall that Quah (1992) could not reject the no convergence hypothesis when applying the random field method to the Summers-Heston data set. This situation contrasts with Barro and Sala-i-Martin (1992a), Mankiw, Romer and Weil (1992) and other writers' findings[3]. There are at least two possible ways of explaining the apparently contradictory findings. Fisrtly, if convergence is conditional, then random field regressions will possibly not detect it. Secondly, if convergence is limited, initial value regressions will usually have a higher probability of producing "convergence-like" results.

3. Identifying the convergence club

3.1 Markov chains methods

Quah (1994) defends that Markov chains methods can be useful to uncover whether polarization is occurring (the richer are getting relatively richer while the poor are getting poorer.) Polarization can be detected in a bimodal long run distribution implied by estimated Markov transition matrices.

It was demonstrated in chapter 4 that a conditional convergence process across *all* economies can generate situations where long run inferred distributions are bimodal. A situation of limited unconditional convergence is analysed here and results using this method with distinct patterns of limited convergence (diverse locations of the convergence club) are considered. In every case, 30 percent of the countries were converging at a speed equal to 0.02. The time

[3] See chapter 2 for developments and references.

length equals 30.

Suppose that the members of the converging club are the 30 percent of the richest countries at the beginning of period. Matrix M_1 is the average of the estimated one-period transition matrices after 1000 replications[4].

$$M_1 = \begin{pmatrix} 0.982 & 0.0175 & 0.0 & 0.0 & 0.0 \\ 0.0176 & 0.974 & 0.0202 & 0.0 & 0.0 \\ 0.0 & 0.00834 & 0.978 & 0.000451 & 0.0 \\ 0.0 & 0.0 & 0.00168 & 0.986 & 0.00412 \\ 0.0 & 0.0 & 0.0 & 0.0134 & 0.996 \end{pmatrix} \quad (6.3)$$

Diagonal elements denote the estimated probability of a country staying within the same quantile. The last element refers to the higher quantile: richer countries almost never swap into a lower quantile. The implied long run distribution of incomes is described by the following vector:

$$F_{1,\infty} = \begin{pmatrix} 0.125 \\ 0.114 \\ 0.0402 \\ 0.148 \\ 0.573 \end{pmatrix} . \quad (6.4)$$

The long run distribution reflects a polarization process going on. The "middle classes" vanish and countries are concentrated either in the highest or in the lowest quantiles.

But what if the convergence club was uniformly dispersed along the sample? In that case, very different results arise. The one-period transition matrix becomes :

[4]See chapter 3 for a complete description of this method. In every case results were derived from the estimation of the one-period transition matrix. Results from the 30-period matrix were qualitatively similar.

$$M_1 = \begin{pmatrix} 0.977 & 0.0143 & 0.0 & 0.0 & 0.0 \\ 0.0233 & 0.960 & 0.0118 & 0.0 & 0.0 \\ 0.0 & 0.0259 & 0.970 & 0.0175 & 0.0 \\ 0.0 & 0.0 & 0.0179 & 0.970 & 0.0176 \\ 0.0 & 0.0 & 0.0 & 0.0129 & 0.982 \end{pmatrix} \quad \textbf{(6.5)}$$

The diagonal elements are now similar and the probability of a country swapping into a neighbouring quantile is small but not negligible. The corresponding long run distribution is:

$$F_{1,\infty} = \begin{pmatrix} 0.0779 \\ 0.125 \\ 0.282 \\ 0.299 \\ 0.215 \end{pmatrix} . \quad \textbf{(6.6)}$$

There is no evidence of a convergence group, and the researcher could possibly conclude that global convergence is at work. Note that the lowest quantile is now the smallest one.

If the convergence group is (unlikely) situated at the end of the sample, the "convergence-like" results would be even stronger. Now, the one-period transition matrix becomes:

$$M_1 = \begin{pmatrix} 0.782 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.218 & 0.958 & 0.0136 & 0.0 & 0.0 \\ 0.0 & 0.0422 & 0.983 & 0.0186 & 0.0 \\ 0.0 & 0.0 & 0.0037 & 0.967 & 0.0203 \\ 0.0 & 0.0 & 0.0 & 0.0203 & 0.980 \end{pmatrix} \quad \textbf{(6.7)}$$

Since the lowest income groups are now converging, the first diagonal element is considerably smaller than one. The interesting thing about this upward mobility is that it results in long run distribution that shows up as convergence:

$$F_{1,\infty} = \begin{pmatrix} 0.0 \\ 0.194 \\ 0.617 \\ 0.114 \\ 0.0747 \end{pmatrix} . \qquad \textbf{(6.8)}$$

Most of the countries are located in the second and third quantiles.

The Markov chains method tends not to detect a limited convergence situation if the converging series are not all at the beginning of the sample. Quah (1994) compares results using the Summers-Heston data set and data for US states. After estimating a bimodal long run distribution, he concludes in favour of a convergence club among the richest countries in the world. In the US case, the unimodal long run distribution is presented as evidence of convergence.

Considering the simulation results, it is possible to argue that Quah's results using country data do not allow for a dismissal of the conditional convergence hypothesis. Moreover, results using US data do not warrant a denial of a limited unconditional convergence hypothesis. If only a part of the states is converging, this could be enough to result in a long run estimated distribution that is unimodal.

This method has another drawback: it does not provide any means of identifying the members of the convergence club.

3.2 Time series methods

When using time series methods it is necessary to use a benchmark series. The test goes on by testing convergence of all the other series to that benchmark. It is not necessary that the benchmark is the real attractor, because "converge to" is transitive. If series A and B converge to series C than series A also converges to series B.

As shown before, time series methods tend to lose power when series start close together[5]. Adopting the "no convergence" hypothesis as the null, there will be cases of converging series that will not show as convergence in the tests. Anyhow, results should be interpreted accordingly. It is both true that Barro-type regressions do not provide overwhelming evidence of convergence across all countries and that time series methods will not allow the researcher to conclude in favour of no convergence as a rule.

Conclusion

This chapter compared different methods of measuring convergence under a data generation process that implied limited convergence among series.

Initial value and random field regressions can erroneously lead the researcher into accepting a "convergence across all series" hypothesis or to dismiss any kind of convergence, even if *some* series are converging. Some convergence findings using the Summers-Heston data base are probably not so informative as one could expect them to be. Namely, some presented simulations allowed for results that are similar to the ones in Barro and Sala-i-Martin (1992a, 1995), in Mankiw, Rommer and Weil (1992) or in Quah (1992).

Alternative Markov chains methods, proposed by Quah (1993a, 1994) also suffer from considerable drawbacks. They do not allow for a conditional convergence process, possibly dismissing convergence then. Moreover, it is possible that these methods display convergence evidence with some patterns of limited convergence. Anyhow, they do not allow the identification of the members of a convergence club.

The advantage of time series methods is to allow for both conditional and unconditional convergence. They depend on the number of periods and on initial conditions. Nevertheless, they do not fail in the presence of a limited convergence situation and belonging of a series to a convergence club can be directly tested.

---

[5] See chapter 3

Chapter 7
Time-varying Convergence

Introduction

Previous chapters considered a number of departures from a "well-behaved" convergence generation process. In this well-behaved process, all series converged to the leading series at a constant rate or speed and once there the long-run differences between each pair of series were equal to zero. Not surprisingly, almost all the methods to measure and test for convergence performed well in this ideal scenario.

In the first main departure, the data generation process implied different steady-state levels for each series. This was called a "conditional convergence" situation and was covered in chapter 4. Chapter 5 included an evaluation of different methods in a framework characterized by the existence of two "convergence clubs": part of the series converge to one of the leading series and the other part converges to a different leader. In a "limited convergence" setup (chapter 6) there was only one leading series but only part of the series converged to it (the other ones wandered around.)

Here, the situation where all the series converge unconditionally to the same leader is reconsidered. All series behave in a similar way and this is a framework similar to the one examined in chapter 3 except for one fundamental difference: convergence starts to take place after some periods of no convergence. This "break period" is made different in various simulations: it can be closer to the beginning or to the end period.

As usual, artificial series are generated that converge according to this time-varying pattern and different testing methods are applied to artificial data.

After a more detailed description of the time-varying convergence generation process (first section) two time series methods (the Kalman filter and Augmented Dickey-Fuller) are assessed and compared in section two. Since time series methods use two series at a time only, they have to be evaluated separately from cross-section methods.

Cross section methods are appraised in the third section. Three methods are considered here: initial value regressions, random field regressions and Markov chains estimations. The data generation process presupposes that 99 series are converging to the same leading series.

The chapter ends with a conclusion, where all the methods are compared and assessed and results are put into perspective taking into account the outcomes from previous chapters.

1. The data generation process

The data generation process is similar to the one used in chapter 3. Series X_1 is the attracting or leading series and generated according to:

$$x_{1,t}=x_{1,t-1}+g+\eta_{1,t}+\Delta\epsilon_{1,t}, \qquad \textbf{(7.1)}$$

where η and ϵ fulfil the roles of permanent and temporary shocks, respectively. These are normally and independently distributed random variables with no autocorrelation. To gain some resemblance to real data, g was made equal to the US GDP per head average growth rate between 1960 and 1989.Similarly, the total variance of the disturbance term is equal to the variance of the US growth rate from 1960 to 1989.

Attracted series (X_2 to X_{100}) are generated from equation (7.2):

$$x_{i,t}=x_{i,t-1}+g+\beta_t(x_{1,t-1}-x_{i,t-1})+\eta_{i,t}+\Delta\epsilon_{i,t}, \qquad i=2,3,\ldots,100. \qquad \textbf{(7.2)}$$

where

$$\beta_t = \begin{cases} \beta, & \text{if } t \geq s, \\ 0, & \text{if } t < s. \end{cases} \qquad \textbf{(7.3)}$$

In equation (7.2) β_t is a time-varying speed of convergence. The fact that β_t equals zero until moment s means that no series is converging to X_1 until that moment is reached. From then on, all series are converging at the same rate β . In (7.2) , η and ϵ are defined as in (7.1).

As in previous chapters, initial values for the series were made equal to the logs of GDPs per head of 100 countries (see the appendix to chapter 3 for a table.) Series are ordered according to these initial values, so that the first series (the attractor) corresponds to the richest country in 1960 (the US) and series 100 to the poorest (Lesotho.)

2. Time series methods

Two methods are comprised in this category: the Kalman filter method and the Augmented Dickey-Fuller one. They were already fully described and used in previous chapters so only a brief reminder is included here. Both methods rely on examining the time series of the differences between one country and the leading country. This difference is from now on defined by $d_{i,t}$:

$$d_{i,t} = x_{1,t} - x_{i,t}. \tag{7.4}$$

2.1 Kalman filter method

The following model is estimated for the time series differences for country i:

$$d_{i,t} = \alpha_{i,t} + \epsilon_{i,t}, \tag{7.5}$$

$$\alpha_{i,t} = \alpha_{i,t-1} + \mu_{i,t}. \tag{7.6}$$

The difference is modelled as a random walk plus noise model. If the variance of μ tends to zero as time tends to infinity, the no convergence hypothesis is dismissed in favour of convergence. If one assumes that $Var(\mu_{i,t}) = \phi_i^2 Var(\mu_{i,t-1})$, an estimate of ϕ that is statistically smaller than one is taken as evidence of convergence.

In the simulations, the number of periods was always 100. Two initial positions for the attracted series were considered: position 50 and position 100. Consequently, the attracted series started with an initial value that is equal either to Malaysia (country number 50) or to Lesotho (country number 100) GDP per head in 1960. Also, two different speeds of convergence were contemplated. β in equation (7.3) was equal either to 0.02 or to 0.10. Finally, different break periods were imposed, this meaning that s in (7.3) took the value of 10, 25 or 50.

Summary results for the simulations are presented in Table 7.1.

Table 7.1
Kalman filter method
Structural break
Percentage of no convergence rejections
5 and 1 percent confidence levels
(1000 replications)

	β=0.02	β=0.10
break=10 initial position=50	25.4 7.9	100.0 99.9
break=10 initial position=100	84.3 50.3	100.0 100.0
break=25 initial position=50	7.1 1.4	95.3 89.6
break=25 initial position=100	11.2 1.7	99.8 99.3
break=50 initial position=50	0.2 0.1	0.2 0.1
break=50 initial position=100	0.0 0.0	0.0 0.0

A clear pattern arises from the analysis of Table 7.1:

- firstly, the number of no convergence rejections augments from left to right; the test is more powerful when the speed is higher. This was already the case for a constant speed of convergence in chapter 3;

- secondly, the percentage of no convergence rejections is smaller when the initial position is equal to 50 (as compared to position 100.) This means that the power of the test is greater when the series start further apart. Again, this happened before in chapter 3;

- finally, it is also clear that the power of the test depends negatively on the break moment. When the moment of the break equals 50, the test is unable to dismiss the no convergence hypothesis.

2.2 Augmented Dickey-Fuller tests

Table 7.2
ADF method
Structural break
Percentage of no convergence rejections
5 and 1 percent confidence levels
(1000 replications)

	β=0.02	β=0.10
break=10 initial position=50	7.5 1.1	99.9 77.7
break=10 initial position=100	42.1 8.8	100.0 95.6
break=25 initial position=50	0.0 0.0	0.0 0.0
break=25 initial position=100	0.0 0.0	0.0 0.0
break=50 initial position=50	0.0 0.0	0.0 0.0
break=50 initial position=100	0.0 0.0	0.0 0.0

This is a unit root test for differences between one series and the leading series. The following equation was estimated by ordinary least squares:

$$d_{i,t}-d_{i,t-1} = b_1 + b_2.d_{i,t-1} + \sum_{j=1}^{l} b_{2+j}.(d_{i,t-j}-d_{i,t-j-1}) + \epsilon_{i,t}. \tag{7.7}$$

l denotes the number of lags of the dependent variable. These are introduced until the residuals do not display any autocorrelation. In this case, they passed a Lagrange multiplier test. If l lags are used, the test is called an ADF(l) test. No convergence (or non-stationarity) is rejected if b_2 is significantly negative. Under the null of no convergence, the standard t-statistic on the b_2 coefficient does not follow a normal distribution, but a Dickey-Fuller one, so that different critical values have to be used. These were provided in chapter 3.

The generated data was the same used in the previous section. Results for the Augmented Dickey-Fuller test are displayed in Table 7.2. The same pattern, already described for the Kalman filter, arises there; power depends negatively on the break period, and positively on the speed of convergence and on the initial distance. Again, the influence of the last two variables is not surprising and was detected before (see chapter 3.)

A comparison of Kalman filter results (Table 7.1) with ADF results (Table 7.2) leads to the conclusion that the first method is more powerful in every case, i. e., more robust in the presence of this type of time-varying speed of convergence.

3. Cross section methods

Three methods are considered under this heading: initial value regressions, random field regressions and Markov chains estimation. They all use every available series. Accordingly, the same artificial data set is used (the one generated according to the procedures specified in the first section.)

3.1 Initial value regressions

Since this is a case of unconditional convergence, there is no reason to introduce any conditioning variables. Equation (7.8) is estimated using OLS across series:

$$\frac{x_{i,T}-x_{i,1}}{T-1} = b_0 + b_1 x_{i,1} + v_{i,T}. \tag{7.8}$$

Here, the dependent variable is the average growth rate and the independent variable is the initial value for each series. The no convergence hypothesis is rejected when there is a statistically significant negative relationship between growth rates and initial values, given by a t-statistic on b_1 that is significantly smaller than zero

As explained before (see chapter 3), an estimate for the speed of convergence is obtained from the following:

$$\hat{\beta} = 1 - [(T-1).\hat{b}_1 + 1]^{\frac{1}{t-1}}, \tag{7.9}$$

where $\hat{b}_1$ is the OLS estimate for b_1. If $(T-1).\hat{b}_1$ is smaller than -1, there is no real estimate for $\hat{\beta}$. This case was registered as a failure in Table 7.3.

Three different positions for the break were considered. Remember that the total number of periods equals 100. In the first situation, convergence starts to take place close to the beginning, so that the break occurs at time 10. In a second hypothesis, the break happens after 50 periods, and finally a break that is close to the end of the interval is considered, making s=90.

Two different speeds of convergence (0.02 and 0.10) were considered for these three different values of s (the break moment) . These correspond to the two last columns in Table 7.3.

1000 replications were run for each of the six different combinations of speeds and break times. Power figures at the 5 and 1 percent confidence levels and the average estimated speed according to equation (7.9) are reproduced in Table 7.3 (the latter figure is written in bold characters.)

Table 7.3
Initial value regressions
Structural break
Equal growth rates
Percentage of no convergence rejections
5 and 1 percent confidence levels
average estimated speed
(1000 replications)

	β=0.02	β=0.10
break=10	100.0 100.0 **0.0182**	100.0 100.0 **0.0584***
break=50	100.0 100.0 **0.0152**	100.0 100.0 **0.0528****
break=90	100.0 100.0 **0.00204**	100.0 100.0 **0.0106**
break=101	4.8 1.4 **0.0000**	4.8 1.4 **0.0000**

*49.1 % of successes.
** 89.0 % of successes.

The fact that the power is always equal to its maximum value (100 per cent) is probably the most striking feature in this table. This happens even when the speed is smaller (equal to 0.02) and the break is close to the last period (s=90.)[1] Clearly this is not accompanied by accurate estimates of the speed of convergence. The average estimated speed is always lower than β . This is due to very distinct factors:

- the estimator is an weighted average of 0 and β . When the break occurs later, the estimated speed is lower;

- when the true speed is higher (equal to 0.1) it often happens that $(t-1).\hat{b}_1$ is smaller

[1]Results in the last row refer to a situation where speed is always zero (the total number of periods equals 100).

than -1, when $\hat{b}_1$ is sufficiently big in absolute terms. These cases are discarded from the computation of the average of $\hat{\beta}$ (but not from the percentage of rejections of the no convergence hypothesis) and bias this average towards zero.

The fact that the mean growth rate (g, in equation (7.2)) was equal for every series from the outset is in part responsible for these very powerful results. If the mean growth rate is equal, the relative position of series when they start converging is not very different from the one in the very beginning, even if they are subject to permanent shocks.

Table 7.4
Initial value regressions
Structural break
Uniformly distributed growth rates
Percentage of no convergence rejections
5 and 1 percent confidence levels
average estimated speed
(1000 replications)

	β=0.02	β=0.10
break=10	100.0 100.0 **0.0184**	100.0 100.0 **0.0584***
break=50	100.0 100.0 **0.0105**	100.0 100.0 **0.0528****
break=70	100.0 98.1 **0.00652**	100.0 100.0 **0.0327**
break=90	24.5 6.7 **0.0026**	100.0 100.0 **0.0120**
break=101	3.1 0.5 **0.0000**	3.1 0.5 **0.0000**

*50.7 % of successes.
** 84.9 % of successes.

In a more realistic setting, mean growth rates are only equal after convergence starts to occur. Before, they are uniformly distributed between -0.02 and 0.07 (approximately the extreme values for the Summers and Heston data set, see Table 4.1, chapter 4):

$$g_i = \begin{cases} g, & if\ t \geq s, \\ U(-0.02,\ 0.07), & if\ t < s. \end{cases} \qquad (7.10)$$

Summary results using this last formulation are presented in Table 7.4. Although still successful, and detecting convergence almost all the times when the break occurs at time 70, power declines sharply after it.

3.2 Random field regressions

Table 7.5
Random field regressions
Structural break
Equal growth rates
Percentage of no convergence rejections
5 and 1 percent confidence levels
average estimated speed
(1000 replications)

	β=0.02	β=0.10
break=10	100.0 100.0 **0.0150**	100.0 100.0 **0.0380**
break=50	100.0 100.0 **0.00639**	100.0 100.0 **0.0100**
break=90	21.2 1.4 **0.00188**	100.0 100.0 **0.00511**
break=101	0.2 0.0 **0.000**	0.2 0.0 **0.000**

As discussed before (see chapter 3) if b_T is the estimator for the regression coefficient of $d_{i,t}$ on its own first lag:

$$b_t = (\sum_{i=2}^{N} \sum_{t=1}^{T} d_{j,t-1}^2)^{-1}(\sum_{i=2}^{N} \sum_{t=1}^{T} d_{j,t}\, d_{j,t-1}), \qquad \textbf{(7.11)}$$

and if d has a unit root, so that $\Delta d_{i,t}=\mu_{i,t}$, it can be shown that, under a set of assumptions[2]:

$$z_f = T\sqrt{\frac{N}{2}}(b_T-1) \rightarrow N(0,1) \qquad as\ T\rightarrow\infty. \qquad \textbf{(7.12)}$$

This was the statistic used here to test for convergence. Since this is a random field regression, and not a time series one, all the differences to the benchmark series (series 1) are used in the estimation of b_T.

Table 7.6
Random field regressions
Structural break
Uniformly distributed growth rates
Percentage of no convergence rejections
5 and 1 percent confidence levels
average estimated speed
(1000 replications)

	β=0.02	β=0.10
break=10	100.0 100.0 **0.0158**	100.0 100.0 **0.0414**
break=50	100.0 100.0 **0.0057**	100.0 100.0 **0.0097**
break=70	81.8 22.6 **0.0029**	100.0 100.0 **0.0065**
break=90	0.0 0.0 **-0.0009**	100.0 89.2 **0.0038**
break=101	0.0 0.0 **-0.0046**	0.0 0.0 **-0.0046**

[2]See Quah (1993d) for more details.

The six combinations of two different speeds and three different break points that were considered in the previous section for the case of equal mean growth rates are repeated here (see Table 7.5.) Results are qualitatively similar to the ones reported for the initial value regressions. Nevertheless, the power for the case of a smaller speed and when the break point is closer to the end is much smaller.

Again, power declines when growth rates before convergence starts are uniformly distributed, as can be inferred from Table 7.6.

3.3 Markov chains

The estimation of Markov chains provided similar results with different speeds of convergence, so only the results when β equals 0.02 are considered here. As before, the number of series and periods is equal to 100. Also, the two cases considered in the last two sections (equal and uniformly distributed growth rates before convergence starts) implied analogous results. The following estimations concern the first case.

As in previous chapters, a grid H implicitly divides the 100 series into five quantiles:

$$H = \begin{pmatrix} 0.25 \\ 0.50 \\ 1.00 \\ 2.00 \end{pmatrix} . \qquad \textbf{(7.13)}$$

For example, the second quantile comprises those countries that have a GDP per head between a quarter and half of the average.

<u>Convergence starts to take place earlier (s=10)</u>

The one-period transition matrix for a simulation corresponding to a case when convergence starts to take place after the 10th moment is presented below:

$$M_1 = \begin{pmatrix} 0.835 & 0.00397 & 0.0 & 0.0 & 0.0 \\ 0.165 & 0.893 & 0.00565 & 0.0 & 0.0 \\ 0.0 & 0.103 & 0.976 & 0.0268 & 0.0 \\ 0.0 & 0.0 & 0.0187 & 0.968 & 0.0980 \\ 0.0 & 0.0 & 0.0 & 0.00512 & 0.902 \end{pmatrix} \qquad \textbf{(7.14)}$$

The estimation process for this matrix was explained in chapter 3. Here, it is remembered that element $m_{i,j}$ denotes the probability of a series that is in the j-th quantile in time t to swap into the i-th quantile in time t+1. The relatively low values in the diagonal (not very close to unity) show that there is mobility across quantiles. For example, a very poor country (one that belongs to the first quantile) has a 16.5 percent probability of changing into the closest quantile in the next period.

The long-run distribution implied by M_1 is expressed by the following vector[3]:

$$F_{1,\infty} = \begin{pmatrix} 0.000732 \\ 0.0305 \\ 0.558 \\ 0.390 \\ 0.0204 \end{pmatrix} . \qquad \textbf{(7.15)}$$

In the long-run, the great majority of countries are concentrated in the third and fourth quantiles. This means that their incomes are higher than one half and less than twice the average. This is taken as evidence of convergence.

Convergence takes place after the 50th period

When s is equal to 50, the one-period transition matrix becomes:

[3]See chapter 3 for an explanation on how to estimate this vector

$$M_1 = \begin{pmatrix} 0.934 & 0.00772 & 0.0 & 0.0 & 0.0 \\ 0.0661 & 0.926 & 0.0141 & 0.0 & 0.0 \\ 0.0 & 0.0664 & 0.976 & 0.0150 & 0.0 \\ 0.0 & 0.0 & 0.00988 & 0.975 & 0.0577 \\ 0.0 & 0.0 & 0.0 & 0.00978 & 0.942 \end{pmatrix} \quad \mathbf{(7.16)}$$

The matrix in (7.16) is different from the matrix in (7.14): diagonal elements are closer to unity, so that mobility across quantiles is more restricted. Accordingly, the long-run underlying distribution displays a slightly different pattern:

$$F_{1,\infty} = \begin{pmatrix} 0.0124 \\ 0.106 \\ 0.498 \\ 0.328 \\ 0.0555 \end{pmatrix} . \quad \mathbf{(7.17)}$$

Now, the percentage of countries that fall into the third and fourth quantiles is high, but smaller than the one reported for the case when convergence started earlier.

Convergence takes place later (s=90)

In this extreme case, some of the diagonal elements of the transition matrix are virtually equal to one:

$$M_1 = \begin{pmatrix} 0.965 & 0.0352 & 0.0 & 0.0 & 0.0 \\ 0.0120 & 0.966 & 0.0224 & 0.0 & 0.0 \\ 0.0 & 0.0209 & 0.972 & 0.00712 & 0.0 \\ 0.0 & 0.0 & 0.0132 & 0.984 & 0.00309 \\ 0.0 & 0.0 & 0.0 & 0.00275 & 0.997 \end{pmatrix} \quad \mathbf{(7.18)}$$

The small probability attached to the swapping of quantiles is reflected in the following distribution vector:

$$F_{1,\infty} = \begin{pmatrix} 0.0960 \\ 0.281 \\ 0.300 \\ 0.159 \\ 0.163 \end{pmatrix} . \qquad \textbf{(7.19)}$$

The situation described by (7.19) is hardly one of unconditional convergence; even if it is not a bimodal distribution, there is a considerable weight of all the quantiles considered.

Conclusion

A time-varying speed of convergence was contemplated in this chapter. This means that the series start to converge to the leading series only after some initial periods of no convergence. The speed of convergence does not vary across series: at any point in time, it is the same for each series.

The power of time series methods (Kalman filter and Dickey-Fuller) decline when compared to a situation where all the series start to converge from the beginning. Also, these methods are more powerful when the series start to converge earlier in time and when the speed is higher. The Kalman filter method proved to be more robust to a time-varying pattern than the Augmented Dickey-Fuller tests. The latter were unable to dismiss the no convergence hypothesis a single time in a framework where convergence starts to take place after 25 periods (in a total of 100.)

Cross-section methods (initial value and random field regressions, Markov chains estimation) were generally successful in dismissing the no convergence hypothesis. The two regression methods could even dismiss it when the break occurs as late as at moment 70 (in a total of 100 periods.)

In this very particular simulation framework, cross-section methods were more successful than time series methods. The reason for this comes from the fact that there is no cross-section variation. Instead, there is a time inconstancy of convergence. Since cross-section methods use all the cross-section information, they are more robust to this kind of variation.

When the source of variation is cross-sectional, as in the previous chapters that included conditional convergence (chapter 4), convergence clubs (chapter 5) and limited convergence (chapter 6) there was case for preferring time series to cross-section ones. Here, the contrary is happening. In practice, the researcher has to balance the advantages in using cross-section methods (using more information) to its drawbacks (imposing cross-sectional restrictions that are not verified by real data.)

Before the empirical application in the next chapter, it is useful to summarise the main conclusions from the Monte Carlo simulations (chapters 3 to 7):

(1) The information content of cross-section dispersion measures is limited. Qualitatively different time paths for these type of measures can arise from situations where all the series converge.

(2) Cross-section methods (initial value and random fields regressions, Markov chains method)
may be misguiding when there is more than one convergence club or if convergence is limited.

(3) Cross-section methods may also give misleading information when convergence is conditional.

(4) Time series methods (Dickey-Fuller and the Kalman filter) are robust to the kind of "cross-section variation mentioned in (2) and can cope with conditional convergence.

(5) Time series methods have more power when the initial distance between series relative to the steady-state is higher. Similarly, they are more powerful when the speed of convergence is greater.

(6) Within time series methods, the Kalman filter one is more powerful in situations where convergence starts only after a few periods of no convergence.

(7) Cross-section methods may produce meaningful information when time is the only source of variation, as it was the case in this chapter.

Chapter 8

Convergence Across Industrialised Countries (1890-1989)

Introduction

In this empirical study, some of the methods previously discussed and evaluated are applied to long series for annual GDP per head for 16 industrialised countries from 1890 to 1989. The data set comprises the following countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and the United States.

The same set of countries was used before by Baumol (1986). In this influential paper, a regression of average growth rates from 1870 to 1979 on initial incomes produced a significant negative value. The author concluded that convergence was at work, and observed that results were different when additional countries were included.

Bernard and Durlauf (1991) considered a long series of GDP per head from 1900 to 1987 for all the aforementioned countries except Switzerland. They tested for the stationarity of the pairwise differences in GDP per head and concluded that there was a set of common factors that jointly determined output growth, but little evidence of convergence.

Other researchers have used data sets that are not very different from this one in what concerns included countries, but considered shorter time spans. Dowrick and Nguyen (1989) statistically compared growth rates for poorer and richer countries in the OECD in different periods from 1950 to 1985 and concluded in favour of convergence. An initial value cross-section regression including some conditioning variables reinforced their conclusion. Barro and Sala-i-Martin (1992a) and Mankiw, Romer and Weil (1992) considered 20 and 22 OECD countries, respectively, in both cases from 1960 to 1985. Also using initial value cross-section regressions, they concluded in favour of conditional convergence.

It might seem that time series tests are in contradiction to cross-sectional ones, the former not allowing the researcher to decide in favour of convergence. Considering the results contained in the chapters 3 to 7, it is likely that one of the following is true:

- industrialised countries are converging but cointegration-type tests fail to recognise that because they have low power, specially in the presence of structural breaks;

-only a part of the considered countries is converging. This is enough for cross-sectional initial value regressions to display a significant negative coefficient for the initial value, and would further explain the failure of a number of pair-wise cointegration tests.

Two time series methods (ADF cointegration tests and a Kalman filter test) are used here in testing for convergence across industrialised countries. These methods imply the choice of a benchmark country and the US seemed to be the natural choice, being the largest economy and the one that generally had the highest GDP per head during the whole period. Accordingly, the convergence tests are tests of convergence of a specific country towards the US.

From the simulation results included in chapters 3 and 7, there was ground to expect that the Kalman filter test would prove to be more powerful when applied to a situation that is likely to be characterized by structural change: the long time interval includes two world wars and a couple of major economic crises.

When the no convergence hypothesis is rejected, the analysis goes further and estimates for both the speed of convergence and the steady-state levels are provided.

The structure of this chapter is as follows.

The first section includes a description of the data set.

Convergence tests to the US level (both Dickey-Fuller and Kalman filter) are presented and discussed in the second section. These tests considered the fifteen series of differences to the US for the whole time period (1890-1989) and also two different subperiods (1890-1939 and 1947-1989) for the restricted group of the G-7 countries.

The third section comprises estimates for the steady-state levels and speeds of convergence for the G-7 countries. Hypotheses that speeds and steady-state levels are the same in the two different subperiods and across countries are tested.

In the fourth and last section a reformulation of the cross-section initial value regressions procedure is presented and its results discussed.

The chapter ends with a conclusion.

All simulation results and empirical estimates included in the chapter were programmed by the author in Gauss and are available on request.

1. The data set

As mentioned above, real annual GDP per head data for 16 industrialised countries from 1890 to 1989 are used in this study. Data are listed in the appendix which also includes graphs plotting the differences in the logarithms of GDP per head between the G-7 countries and the US.

Except for Germany, GDP figures at 1985 US relative prices from 1890 to 1989 were computed after tables presented by Maddison (1991). Population figures came from the same source for years from 1890 to 1949 and from Summers and Heston (Penn World Tables, Mark 6, NBER gopher) for years from 1950 to 1989.

Maddison (1979) is the source for Germany GDP per head from 1890 to 1977. The figures from 1978 to 1989 came from the above mentioned sources. For all countries, series correspond to 1989 borders[1]. Finally, note that German GDP per head is the GDP per head of West Germany.

---

[1]Maddison (1989) includes GDP data adjusted for territorial change and population figures not adjusted for it. Since figures referring to population changes in years of changing borders are also given, population series were adjusted assuming that population in gained or lost territories grew at the same rate as in the rest of the country.

2. Convergence tests

2.1 Dickey-Fuller and Augmented Dickey-Fuller tests

Dickey-Fuller tests for convergence are stationarity tests for the differences between each country GDP and the US GDP.

Equation (8.1) was estimated using OLS. d_t is the difference between the country which is being discussed and the US GDPs per head.

$$d_t - d_{t-1} = b_1 + b_2 . d_{t-1} + \sum_{j=1}^{l} b_{2+j} . (d_{t-j} - d_{t-j-1}) + \epsilon_t . \qquad \textbf{(8.1)}$$

The number of lags denoted by l was chosen so that residuals would `not display significant autocorrelation or non normality[2]. A Lagrange multiplier test for autocorrelation and the Bera-Jarque normality test were used for this effect[3]. In some cases, some dummy variables were introduced in equation (8.1) to take account of wars and economic crises. These dummy variables are of the form:

$$DU_{yt} = \begin{cases} 1, & \textit{if } t=y, \\ 0, & \textit{otherwise}. \end{cases} \qquad \textbf{(8.2)}$$

Usually, the dummies were introduced to account for the big shocks that correspond to the two world wars years; they are listed by country in Table 8.13 (in the appendix.)

If equation (8.1) does not include lagged values of Δd_t, the test is called a "Dickey-Fuller" (DF) test. If l is one or greater, the test becomes an "Augmented Dickey-Fuller" (ADF) one. In both cases the statistic of concern is the t-statistic for the b_2 coefficient. Under the null hypothesis of non-stationary, this statistic is not normally distributed. The relevant critical

[2] In some cases there was no need to introduce the lagged dependent variable in the right-hand side of the equation.

[3] See Cuthbertson, Hall and Taylor (1992) for a presentation of these tests.

Table 8.1
Critical values for the DF and ADF t-statistic
5 % and 1% significance levels

significance level	**t=43**	**t=50**	**t=100**
5 per cent	-2.93	-2.92	-2.89
1 per cent	-3.59	-3.58	-3.50

Source: MacKinnon (1991).

values are reproduced in Table 8.1[4].

Three different periods were considered: the period that goes from 1890 to 1989 and two subperiods chosen to exclude the Second World War. The first subperiod starts in 1890 and ends in 1939 and the second one goes from 1947 to 1989. The countries that constitute the "G-7 group" (Canada, France, Germany, Italy, Japan, the United Kingdom and the United States) were analysed in more detail. They were the only ones considered in the two sub-periods[5].

Dickey-Fuller tests results are summarised in Table 8.2. Values within brackets correspond to the number of lags (given by *l*) in equation (8.1).

More often than not, the "no convergence" hypothesis can not be dismissed. Considering the period from 1890 to 1989, this hypothesis is rejected for five countries (in sixteen) at the five or one percent confidence level. The five countries are: Australia, Belgium, France, the Netherlands and Switzerland. Among the G-7 countries, and considering the period before the Second World War, "no convergence" is only dismissed in the French and German cases. In the post-war period, the null hypothesis is rejected for Germany, Italy, and Japan, but not for France, Canada or the United Kingdom.

[4]Values were taken form MacKinnon(1991).

[5]The US is not included in the table because differences in respect to that country GDP are considered.

Table 8.2
Dickey-Fuller convergence to the US tests

Country	1890-89	1890-1939	1947-1989
Canada	-0.5943 (3)	-2.404 (0)	-0.3441 (0)
France	-4.060** (0)	-3.980** (0)	-1.930 (2)
Germany	-2.720 (1)	-2.630 (0) -3.433** (0) until 1937	-3.709** (1)
Italy	-2.359 (2)	-2.036 (0)	-3.230** (0)
Japan	-0.7851 (0)	-1.104 (0)	-3.597** (0)
United Kingdom	-2.316 (0)	-1.930 (0)	-1.838 (0)
Australia	-3.063(0)*		
Austria	-2.640(1)		
Belgium	-5.266**(4)		
Denmark	-2.602(1)		
Finland	-0.4153(0)		
Netherlands	-3.301* (0) from 1900		
Norway	-0.780 (1)		
Sweden	-0.809 (1)		
Switzerland	-3.112* (1)		

* "No convergence" rejected at the 5 per cent confidence level;
** "No convergence" rejected at the 1 per cent confidence level.

2.2 Kalman filter tests

a) The method

The Kalman filter test used in this chapter was described in more detail in Chapter 3 and is a version of a convergence test proposed by Hall, Robertson and Wickens (1993). Recall the following formulation, where d_t is the difference between the series:

$$d_t = \alpha_t + \epsilon_t, \qquad \textbf{(8.3)}$$

$$\alpha_t = \alpha_{t-1} + \mu_t, \tag{8.4}$$

$$\epsilon_t \sim N(0,\sigma^2), \tag{8.5}$$

$$\mu_t \sim N(0,\Omega_t), \tag{8.6}$$

$$\Omega_t = \phi^2 \Omega_{t-1}, \tag{8.7}$$

$$\Omega_0 = \Upsilon^2. \tag{8.8}$$

Although the noise variance (σ^2) is constant through time, the variance of μ may display a declining pattern. If ϕ is less than 1, this variance tends to 0 in the long run, meaning

Table 8.3
Critical values for the $T(\phi_{ML})$ statistic

%	t-value
0.5	-3.702
1	-3.479
5	-2.479
10	-1.970
50	-0.059
99	3.348
99.5	3.529

that the two series are converging, their difference becoming a stationary variable.

In this model written in state-space form equation (8.3) is the measurement equation and equation (8.4) is the state one. The likelihood function can be constructed using the Kalman filter. The maximum likelihood (ML) estimates for the parameters are obtained maximizing it.

The null hypothesis of no convergence implies that $\phi=1$. Therefore, the proper test for convergence is: H_0: $\phi=1$ against H_a: $\phi<1$. The implementation of this test requires the distribution of (some function of) ϕ under the null. The distribution of the following was tabulated after 1000 replications:

$$T(\phi_{ML}) = \frac{\phi_{ML} - 1}{\sqrt{(h^{-1})_{22}}}. \qquad \textbf{(8.9)}$$

In equation (8.9), $(h^{-1})_{22}$ is the second element in the diagonal of the inverse of Hessian matrix. The tabulation is reproduced in Table 8.4.

b) The results

Table 8.4 displays the results for the $T(\phi_{ML})$ statistic using the Maddison data. Again, some dummies of the form given by (8.2) were introduced in the measurement equation[6].

For the longer period, the no convergence hypothesis could be dismissed for all countries considered, at least at the 5 % confidence level. This was not the case for the Dickey-Fuller test. It was shown in chapter 5 that the Kalman filter test tended to be more powerful when convergence does not work from the very beginning of the sample period. This can well be the case here.

[6] The residuals for Japan and Canada in the period from 1947 to 1989 did not pass the normality test.

On the other hand, Dickey-Fuller tests seem to be more powerful in dismissing convergence

Table 8.4
Kalman filter convergence to the US tests

Country	1890-1989	1890-1939	1947-1989
Canada	-7.212**	1.035	-1.710
France	-7.983**	-0.300	-4.693**
Germany	-3.723**	0.709	-6.469** (from 1950)
Italy	-6.432**	-0.391	-5.292**
Japan	-3.640**	3.010	-4.108**
United Kingdom	-6.464**	0.120	-3.090*
Australia	-9.184**		
Austria	-3.241*		
Belgium	-6.162**		
Denmark	-5.967**		
Finland	-3.656**		
Netherlands	-7.603** (from 1900)		
Norway	-3.650**		
Sweden	-4.471**		
Switzerland	-4.078** (from 1899)		

* "No convergence" rejected at the 5 per cent confidence level;
** "No convergence" rejected at the 1 per cent confidence level.

when the first sub-period is considered. With this second method, the no convergence hypothesis was never dismissed. According to the simulation results described in chapter 3, this could be due to the fact that the countries considered were already close to their steady-state in 1890. Results for the post-war period are stronger: no convergence is rejected for all the G-7 countries, except for Canada. This last result is probably due to the fact that Canada was the country closest to the US after the Second World War (time series methods have low power when the difference between two series starts close to the steady state.)

2.3 Summary of results

It can be concluded that the sixteen countries under analysis were converging from 1890 to 1989. The post-war period (1947-1989) was probably the time where that convergence process was more evident. Nevertheless, the failure to detect convergence in the early period (1890-1989) can be due to two reasons:

- countries started converging somewhere between 1890 and 1939;
- countries were already close to their steady-state in the beginning of the period.

For France and Germany it was possible to reject the no convergence hypothesis even in the first sub-period.

It must be stressed that both ADF and the Kalman filter methods test for conditional convergence[7], so that countries are not necessarily converging to the same level of income. Take the case of France, for example. The no convergence hypothesis was rejected for the whole period and for each subperiod but Figure 8.4 (in the appendix) suggests that the steady-state level of French income is not equal to the US one. Before the Second World War, and excluding the First War and the 1929 crisis, there was no asymptotic approximation towards the US level. After the forties, and following the big Second World War shock, there is a marked tendency to an approximation to US levels. Nevertheless, this approximation seems to have stopped in the seventies, suggesting that the steady-state level was attained by then.

In the following section, estimates for the different steady-states across countries are provided. Also, the hypothesis that the steady-state is different in the two subperiods is considered and tested.

[7] See chapter 1 for different definitions of convergence.

3. Estimating the speed of convergence and the steady-state

3.1 Estimates for the speed of convergence and the steady-state using OLS

Consider again equation (8.1), reproduced below for your convenience.

$$d_t - d_{t-1} = b_1 + b_2.d_{t-1} + \sum_{j=1}^{l} b_{2+j}.(d_{t-j} - d_{t-j-1}) + \epsilon_t. \qquad \textbf{(8.10)}$$

The steady-state difference between the two countries GDP is given by d_{ss}:

$$d_{ss} = -\frac{b_1}{b_2}. \qquad \textbf{(8.11)}$$

This means that an estimator of the steady-state is given by a function (a ratio) of two estimated parameters:

$$\hat{d}_{ss} = -\frac{\hat{b}_1}{\hat{b}_2}. \qquad \textbf{(8.12)}$$

An estimate for the variance of this estimator is given by[8]:

$$\hat{Var}(\hat{d}_{ss}) = \begin{bmatrix} \frac{\partial d_{ss}}{\partial b_1} & \frac{\partial d_{ss}}{\partial b_2} \end{bmatrix} . \hat{Cov}(\hat{b}_1, \hat{b}_2) . \begin{bmatrix} \frac{\partial d_{ss}}{\partial b_1} \\ \frac{\partial d_{ss}}{\partial b_2} \end{bmatrix}, \qquad \textbf{(8.13)}$$

where the derivatives are evaluated at the point $(\hat{b}_1, \hat{b}_2)$ and $\hat{Cov}(\hat{b}_1, \hat{b}_2)$ is the estimated variance-covariance matrix of $(\hat{b}_1, \hat{b}_2)$.

[8]See Breusch and Wickens (1988).

The annual speed of convergence (call it β) can be estimated from $\hat{b}_2$:

$$\hat{\beta} = 1 - \hat{b}_2. \qquad \textbf{(8.14)}$$

The difference between the two countries GDPs approaches its steady-state value at a rate approximately equal to β percent per year.

3.2 Estimates for the speed of convergence and the steady-state using the Kalman filter

Consider the model described by equations (8.15) and (8.16):

$$d_t = d_{ss} + b_t.(d_{t-1} - d_{ss}) + \epsilon_t, \qquad \textbf{(8.15)}$$

$$b_t - \bar{b} = a(b_{t-1} - \bar{b}) + \mu_t, \qquad \textbf{(8.16)}$$

where a is between -1 and 1. ϵ_t and μ_t are i.i.d. normal disturbances.

b_t is a time-varying speed of convergence. Nevertheless, and considering that |a|<1, it tends to return to its "long run" value, given by $\bar{b}$. This is a case of a "return to normality" model.[9]

Equations (8.15) and (8.16) are the measurement and transition equations in a state-space representation of the return to normality model. The concentrated log-likelihood can be written as an implicit function of the relevant hyperparameters and maximized using numerical methods. Here, the estimated hyperparameters are the steady-state level d_{ss}, the long run speed of convergence $\bar{b}$, the autoregressive parameter for the short-run speed of convergence a, and the variance in the transition equation (as a ratio to the variance in the measurement equation.)

[9] See Harvey (1989).

The inverse of the Hessian matrix is an estimate for the variance-covariance matrix of the hyperparameters and was used in the construction of the t-tests presented below.

3.3 The period from 1890 to 1989

Estimates for the steady-state and the speed of convergence using both OLS and the Kalman filter are presented in Table 8.5[10]. Numbers inside brackets are t-statistics.

These estimates were not very different according to the method of estimation, specially when they were statistically significant. It appears that the state-space formulation is a more parsimonious way of representing a process that is basically the same.

The speed of convergence and the steady-state estimates were not statistically significant in two cases: Japan and Canada. One possible interpretation can be applied to both countries. If convergence started after 1890 it is very likely that Dickey-Fuller-type tests do not detect it. Moreover, and according to earlier simulation results, it is very likely in that case that convergence is detected by a Kalman filter test. This was the case for both countries: from Tables 8.2 and 8.4 it can be noted that Dickey-Fuller t-statistics were not significant while the $T(\phi_{ML})$ statistic was significantly smaller than zero in both cases.

France, Germany, Italy and the United Kingdom all display speeds that are higher than zero. Their steady-states are all significantly smaller than zero. This shows that these countries were converging conditionally towards the United States. In the long run, the difference between the logs of incomes per head is a stationary variable with a non-zero mean.

[10] The t-statistics in the first column of the table are equal to the ones presented before as "Dickey-Fuller tests".

Table 8.5
Speed of convergence and steady-state
OLS and Kalman filter estimates
1890-1989

	speed (OLS)	speed (KF)	steady-state (OLS)	steady-state (KF)	KF log likelihood
Canada	0.0155 (-0.594)	0.004 (-0.397)	0.136 (0.160)	1.403 (0.347)	345.5
France	0.127 (-4.060)	0.127 (-6.010)	-0.420 (-8.507)	-0.419 (-12.41)	291.6
Germany	0.072 (-2.720)	0.077 (-2.918)	-0.479 (-5.437)	-0.400 (-6.739)	292.1
Italy	0.052 (-2.359)	0.052 (-2.315)	-0.458 (-2.740)	-0.394 (-4.321)	300.3
Japan	0.013 (-0.785)	0.005 (-0.602)	0.531 (0.245)	2.420 (0.490)	278.7
United Kingdom	0.078 (-2.316)	0.069 (-2.532)	-0.310 (-4.105)	-0.298 (-4.190)	284.6

Incomes per head were taken in log form. Let Y^s_{us} and Y^s_i denote the steady-state levels for the US and country i incomes, taken in original units. Since:

$$\ln(Y_i^s) - \ln(Y_{us}^s) = ss_i \Rightarrow \frac{Y_i^s}{Y_{us}^s} = \exp(ss_i), \qquad \textbf{(8.17)}$$

it is possible to express the steady-state values presented in Table 8.5 as a percentage of US GDP per head. This is done in Table 8.6, for the Kalman filter estimates only.

Table 8.6
Steady-state as a percentage of the US
1890-1989

	steady-state (% of US)
France	69.8
Germany	61.8
Italy	63.3
United Kingdom	73.3

3.4 Are the estimates stable through time?

The Second World War was a tremendous shock to the G-7 GDPs, as can be seen directly from the graphs. As mentioned before, it was necessary to introduce some dummies to account for this shock and bring the residuals back into normality.

These shocks and the data themselves suggest that different patterns of convergence could have existed before and after the Second World War[11].

To test for stability, the sample was divided into two parts. The first one goes from 1890 to 1946. The second from 1947 to 1989.

First, an unrestricted model that allows for different parameters in the two subperiods was estimated using the Kalman filter. This model is described by the following equations:

$$d_t = d_{ss} + \Delta d_{ss}.D_{1947} + b_t.(d_{t-1} - (d_{ss} + \Delta d_{ss}.D_{1947})) + \epsilon_t, \qquad \textbf{(8.18)}$$

$$b_t - (\bar{b} + \Delta\bar{b}.D_{1947}) = (a + \Delta a.D_{1947})(b_{t-1} - (\bar{b} + \Delta\bar{b}.D_{1947})) + \mu_t, \qquad \textbf{(8.19)}$$

[11]This is a hypothesis that Maddison (1989) defends.

$$Var(\mu_t) = (Q + \Delta Q . D_{1947}) . Var(\epsilon_t). \tag{8.20}$$

$$DU_{1947} = \begin{cases} 1, & if\ t > 1946, \\ 0, & otherwise. \end{cases} \tag{8.21}$$

There are eight hyperparameters in this unrestricted model:

- the steady-state in the period 1890-1946;
- the change in the steady-state from the first period to the second one (1947-1989);
- the speed of convergence in the first period;
- the change in speed from the first to the second period;
- the autoregressive parameter in the transition equation in the first period;
- the change in the autoregressive parameter;
- the variance in the transition equation (as a proportion of the variance in the measurement equation);
- the change in the variance in the transition equation.

The restricted model determines that four of these parameters are equal to zero (the changes), so that speeds, steady-states and autoregressive parameters are the same in both periods. This is the model described by equations (8.15) and (8.16) and its results were already discussed. The same dummies were introduced in the restricted and unrestricted models (see Table 8.15, in the appendix.)

Table 8.7
Return to normality model
Unrestricted version
1890-1989

	steady-state 1890-1946	change in steady-state 1947-1989	convergence speed 1890-1946	change in speed 1947-1989	log likelihood
Canada	1.193 (0.622)	-0.13 (-0.070	0.005 (0.9660	-0.00 (-0.143)	355.7
France	-0.53 (-28.00)	0.29 (4.40)	0.293 (5.79)	-0.219 (-4.147)	305.1
Germany	-0.54 (-9.76)	0.28 (4.689)	0.112 (3.627)	-0.003 (0.10)	310.3
Italy	-0.83 (-18.58)	0.56 (5.219)	0.119 (2.407)	-0.06 (1.138)	308.2
Japan	-1.34 (-30.46)	1.44 (6.574)	0.166 (4.733)	-0.122 (-3.209)	302.8
United Kingdom	-0.33 (-3.983)	-0.018 (-0.200)	0.085 (3.5530	0.069 (0.78)	299.1

A likelihood ratio test was performed to check the validity of the four restrictions. If the restrictions are valid, the following statistic is asymptotically distributed as a chi-square with four degrees of freedom:

$$LR = 2.(\ln(L_u) - \ln(L_r)). \qquad \textbf{(8.22)}$$

In equation (8.22) $\ln(L_u)$ is the natural logarithm of the maximum likelihood for the unrestricted model and $\ln(L_r)$ is the same quantity for the restricted model.

Table 8.8
Steady-state as a percentage of the US
1890-1946 and 1947-1989

	steady-states:	
	1890-1946	1947-1989
France	58.9	78.7
Germany	58.3	77.1
Italy	43.6	76.3
Japan	26.2	110.5
United Kingdom	71.9	70.6

Table 8.7 displays some results for the unrestricted model. Values within brackets are t-values.

In all cases the hypothesis of no structural change was rejected at less than one percent significance level. With the exception of Canada, the speed of convergence was always found to be positive.

The two steady-states are compared in Table 8.8 and presented as a percentage of US GDP (Canada was excluded.)

For France, Germany, Italy and Japan there was a significant increase in the steady-state. The change for this last country was dramatic: the steady-state for the post-war period is estimated to be above the US level. The United Kingdom was the only country for which the steady-state did not increase. It was also the country with the highest steady-state in the pre-war period among the five considered in the table.

The speed of convergence suffered significant decreases in France and Japan. In all the other countries the corresponding t-values were not sufficiently high in absolute value in order to dismiss the no change hypothesis.

3.5 Convergence after the Second World War

In the last section some estimates using the Kalman filter and referring to the after war period were presented. Here, results using OLS are discussed and two hypothesis are examined. First, the equality of speeds, and second, the equality of steady-states across countries are tested.

Both tests are likelihood ratio tests. This kind of tests always imply the estimation of both the

Table 8.9
Speed of convergence estimates
1950-1989

	unrestricted version	**restricted version**
France	0.091 (3.441)	0.0641 (8.563)
Germany	0.136 (4.283)	0.0641 (8.563)
Italy	0.086 (3.464)	0.0641 (8.563)
Japan	0.033 (2.461)	0.0641 (8.563)
United Kingdom	0.319 (3.578)	0.0641 (8.563)

restricted and unrestricted models.

The period from 1950 to 1989 was considered. The following equation describes the general model:

$$d_{it}-d_{i,t-1} = b_{i1}+b_{i2}.d_{i,t-1}+\sum_{j=1}^{3} b_{i,2+j}.(d_{i,t-j}-d_{i,t-j-1})+\epsilon_{i,t}. \qquad \textbf{(8.23)}$$

Five countries were contemplated: France, Germany, Italy, Japan and the United Kingdom, and, as before, differences from the US GDP were taken. Canada was excluded because the unit root hypothesis could not be dismissed.

The number of lags in equation (8.23) was sufficiently high to allow the residuals not to display any correlation and to pass the Bera and Jarque normality test in each of the considered series.

The unrestricted model was estimated by maximum likelihood[12]. The log-likelihood function to be maximised is:

$$\log L(b,\sigma) = -\frac{1}{2}\log(2\pi)-\frac{1}{2}\log|\Theta| -\frac{1}{2}\sum_{i=1}^{5} \left\{(Y_i-X_i'b_i)\Theta_i^{-1}(Y_i-X_i'b_i)\right\}, \qquad \textbf{(8.24)}$$

where

$$Y_i = \begin{pmatrix} d_{i5}-d_{i4} \\ \dots \\ d_{iT}-d_{i,T-1} \end{pmatrix}, \qquad X_i = \begin{pmatrix} 1 & d_{i4} & d_{i4}-d_{i3} & \dots & d_{i2}-d_{i1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & d_{iT} & d_{i,T-1}-d_{i,T-2} & \dots & d_{i,T-3}-d_{i,T-2} \end{pmatrix},$$

$$\sigma = \begin{pmatrix} \sigma_1 \\ \dots \\ \sigma_5 \end{pmatrix}, \qquad b_i = \begin{pmatrix} b_{i1} \\ \dots \\ b_{i5} \end{pmatrix}, \qquad \Theta_i = \sigma_i^2.I, \qquad b = \begin{pmatrix} b_1 \\ \dots \\ b_5 \end{pmatrix}.$$

The equality of speeds restriction is expressed as follows:

$$b_{12}=b_{22}=\dots=b_{52}=b_{R2}. \qquad \textbf{(8.25)}$$

The corresponding restricted log-likelihood function is expressed as in (8.24) but this time with:

[12]This is equivalent to the estimation of each of the five single equations using OLS.

$$b_i = \begin{pmatrix} b_{i1} \\ b_{R2} \\ \dots \\ b_{i5} \end{pmatrix}, \qquad b_{NR} = (b_{11} \;\; b_{13} \;\; b_{14} \;\; b_{15} \mid \cdots \mid b_{51} \;\; b_{53} \;\; b_{54} \;\; b_{55})', \qquad b = \begin{pmatrix} b_{R2} \\ - \\ b_{NR} \end{pmatrix}.$$

Table 8.9 contains estimates for the restricted and unrestricted speeds. Values within brackets are t-values and were computed from the inverse of the Hessian matrix.

Table 8.9 suggests that speeds are different across countries. The value for Britain is ten times greater than the value for Japan. In fact, this is the conclusion drawn from the likelihood test, summarised in Table 8.10. Nevertheless, Japan is converging to a steady-state that is higher than the British one (and possibly higher than the US one.) Supposing that British GDP is converging at a high rate to its steady-state (32 percent per year); approximately a third of a shock that deviates it from its long run value as a percentage of US GDP is eliminated within one year. In practice, that means that British GDP per head is always not very far from its steady-state level (estimated as 71.1 percent of the US value.)

Table 8.10
Log-likelihood tests
Same speed and same steady-states across countries
1950-1989

unrestricted log likelihood	restricted log likelihood (same speed)	χ^2 statistic significance level (same speed)	restricted log-likelihood (same steady-state)	χ^2 statistic significance level (same steady-state)
679.80	667.75	0.0	674.88	0.080

The fact that Japan is converging to a higher steady-state (estimated to be 104.1 percent of US GDP) at a lower rate (3.3 percent per year) implies that Japanese growth rates are generally higher than US ones. This happens because Japan is starting from a lower level than the US. Once the Japanese steady-state is reached, one can expect Japan growth rates to become similar to the US ones.

The LR statistic for the equality of speeds test takes the value 24.1 (=2.(679.8-667.75), see Table 8.10.) Its p-value is virtually zero.

Consider now the equality of steady-states hypothesis. This restriction is given by:

$$\frac{b_{11}}{b_{12}} = \frac{b_{21}}{b_{22}} = \ldots \frac{b_{51}}{b_{52}} = b_s. \tag{8.26}$$

Again the corresponding restricted log-likelihood function is expressed as in (8.24), in this case with:

$$b_i = \begin{pmatrix} b_s . b_{i2} \\ b_{i2} \\ \ldots \\ b_{i5} \end{pmatrix}, \quad b_{NR} = (b_{12} \ \ldots \ b_{15} \ | \ \cdots \ | \ b_{52} \ \ldots \ b_{55})', \quad b = \begin{pmatrix} b_s \\ - \\ b_{NR} \end{pmatrix}.$$

Results from the restricted and unrestricted models are presented in Table 8.11. As usual, values within brackets are t-values.

It becomes apparent that steady-state values are not very different across countries, particularly if Japan is excluded. In Table 8.12 these values are expressed as a percentage of the US GDP per head. French, German, Italian, and British values are all close to three quarters of the American figure.

The formal likelihood test result can be read from Table 8.10. The restricted log-likelihood is sufficiently close to the unrestricted one to allow the non-rejection of the restriction at the eight percent confidence level. If Japan had been excluded from the set of countries this last level would probably be higher.

Table 8.11
Steady-state estimates
1950-1989

	unrestricted version	restricted version
France	-0.246 (-4.023)	-0.302 (-13.529)
Germany	-0.282 (-9.050)	-0.302 (-13.529)
Italy	-0.284 (-3.358)	-0.302 (-13.529)
Japan	0.0405 (0.138)	-0.302 (-13.529)
United Kingdom	-0.341 (-28.78)	-0.302 (-13.529)

Table 8.12
Steady-state estimates as percentage of US
1950-1989

	steady-state
France	78.2
Germany	75.4
Italy	75.3
Japan	104.1
United Kingdom	71.1

4. Initial value regressions reconsidered

It must be remembered that "initial value" or Barro regressions are cross-section regressions of the growth rate on the initial level values. These regressions often include some conditioning variables to proxy for different steady-state levels.

It has been argued before that these regressions are not an adequate test for convergence. Results from this type of regressions are interpreted here *after* performing other convergence tests. Only series that passed the convergence test are included.

The first regression was an "unconditional" regression, a simple regression of growth rates on initial values:

$$\begin{aligned} g_{i,50-89} &= 0.0654 - 0.0232 y_{i,50}, \\ &\quad (18.325)\ (-9.3637) \\ R^2 &= 0.945, \quad \hat{\sigma} = 0.00326, \\ \hat{\beta} &= 0.059. \end{aligned} \qquad \textbf{(8.27)}$$

Six countries were included: France, Germany, Japan, Italy, the United Kingdom and the United States. $g_{i,50\text{-}89}$ is the average growth rate of country i from 1950 to 1989 and $y_{i,50}$ is the GDP per head for country i in 1950. Values within brackets are t-statistics. $\hat{\beta}$ is an estimate of the speed of convergence derived from the coefficient on the initial value.

Equation (8.27) imposes the same steady-state level across the five countries. In the last section, this hypothesis could not be rejected at the eight percent confidence level. This is probably why results are highly significant (t-values are very high in absolute terms, even allowing for the low number of degree of freedom.)

Results can be improved if *estimates* for each country steady-state are included as an explanatory variable, as in the following estimated equation:

$$\begin{aligned} g_{i,50-89} &= 0.0676 - 0.0224 y_{i,50} + 0.0177 \hat{ss}_i \\ &\quad (72.10)\ (-35.77)\ \ (7.81) \\ R^2 &= 0.998, \quad \hat{\sigma} = 0.000815, \\ \hat{\beta} &= 0.052. \end{aligned} \qquad \textbf{(8.28)}$$

The steady-state estimates were taken from Table 8.11 (the US value equals zero).

All the t-statistics are highly significant. As expected, the higher the steady-state the higher the growth rate is.

There are some important differences in the approach followed in this section and the Barro and Sala-i-Martin (1992a, 1992b, 1995) approach. Firstly, initial value regressions are not being used as convergence tests. Secondly, the conditioning variables are not proxies for different steady-states. Estimates for the different steady-states were used as conditioning variables, instead. These estimates were obtained from time series methods.

Conclusion

Convergence of GDP per head to US levels for 15 industrialised countries from 1890 to 1989 was formally tested using two different methods: Dickey-Fuller-type tests and Kalman filter tests. Kalman filter tests were more powerful in dismissing the "no convergence hypothesis". In fact, the last method rejected this hypothesis in every case.

When the period was split into two subperiods of more or less the same length and which excluded the Second World War years (1890-1939 and 1947-1989) results were different. Considering only Canada, France, Germany, Italy, Japan and the United Kingdom, it is possible to see that the no convergence hypothesis was dismissed in every case in the more recent period (except for Canada), but only for Germany and France in the first period. One possible explanation is that convergence started to take place somewhere between 1890 and 1939, was disrupted by the Second World War and resumed soon after it.

On the whole, the Kalman filter results give strength to earlier results using cross-section methods and favouring convergence across industrialised countries as mentioned in the introduction to this chapter. The possibility that tests using cointegration techniques (also used here) did not dismiss the no convergence hypothesis because they were not robust enough in the presence of time-varying convergence is also reinforced.

Some direct attempts were presented to measure the steady-state and the speed of convergence for each of the G-7 countries (excluding the US, always used as a benchmark.) Also, the estimated steady-state levels before and after the war were compared.

Steady-state levels are generally higher after the war, except for the United Kingdom and Canada. Equality of steady-state levels after the war (1950-1989) at a level of roughly three quarters of the US one could not be rejected, even if Japan appears as an outlier. The estimated steady-state for this country was somewhat higher than the US level.

Estimates for the speed of convergence were disparate across countries and time. Equality of speeds in the post-war period was rejected and estimated values varied from 0.03 (Japan) to 0.319 (United Kingdom.) A correct interpretation of these figures has to take the different steady-state levels into account.

Appendix to chapter 8

Table 8.13
Convergence to the US
Dickey-Fuller tests
Years with dummy variables

Country	1890-89	1890-1939	1947-1989
Canada	1918-19, 1931, 1943, 1946-47	none	1951, 1956, 1974-75
France	1918, 1940-46	1918	none
Germany	1923-24, 1945-46	1919, 1923	none
Italy	1897, 1919, 1941-46	none	none
Japan	1942-46	none	none
United Kingdom	none	none	none
Australia	none		
Austria	1917-19, 1940-45		
Belgium	1918-19, 1931, 1943, 1946-47		
Denmark	1914, 1918, 1940-46		
Finland	1918-19, 1940-46		
Netherlands	1918-19, 1940-46		
Norway	1932, 1946		
Sweden	1946		
Switzerland	none		

Table 8.14
Convergence to the US
Kalman filter tests
Years with dummy variables

Country	1890-89	1890-1939	1947-1989
Canada	1946-47	none	1951, 1956, 1974-1976
France	1940-45	1918	1949, 1970, 1982, 1984
Germany	1923, 1940-45	1923	1970, 1982
Italy	1941-46	none	1974, 1980, 1982, 1984
Japan	1932, 1942-46	1894, 1932	none
United Kingdom	1932, 1946	none	none
Australia	1946		
Austria	1945-46, 1948-49		
Belgium	1918, 1940-45		
Denmark	1940-45		
Finland	1918-19, 1942, 1946		
Netherlands	1940-46		
Norway	1930, 1932, 1940-46		
Sweden	1946-47		
Switzerland	1941-47		

Table 8.15
Dummies in the return to normality model

	dummies in:
Canada	1918-19, 1931, 1943, 1946-47
France	1918, 1940-46
Germany	1923, 1940-46
Italy	1897, 1919, 1936, 1941-46
Japan	1932, 1938, 1942-46
United Kingdom	none

Table 8.16

GDP per head in the G-7 countries (1870-1989)

(1985 US 1000 relative dollars)

Year	Canada	France	Germany	Italy	Japan	UK	U.S.A.
1870	1.330	1.571	1.300	1.214	0.618	2.682	2.254
1871	1.357	1.604	1.285	1.219		2.812	2.299
1872	1.326	1.778	1.368	1.193		2.790	2.242
1873	1.421	1.657	1.414	1.232		2.828	2.388
1874	1.431	1.854	1.504	1.224		2.846	2.322
1875	1.383	1.902	1.494	1.254		2.886	2.384
1876	1.284	1.774	1.468	1.218		2.883	2.361
1877	1.345	1.825	1.441	1.208		2.879	2.380
1878	1.288	1.793	1.492	1.216		2.860	2.427
1879	1.395	1.693	1.439	1.224		2.818	2.674
1880	1.435	1.803	1.412	1.277		2.923	2.929
1881	1.606	1.878	1.436	1.185		2.999	2.957
1882	1.648	1.946	1.451	1.279		3.062	3.065
1883	1.628	1.919	1.521	1.265		3.062	3.062
1884	1.745	1.883	1.545	1.264		3.044	3.051
1885	1.626	1.852	1.572	1.277	0.683	3.004	2.999
1886	1.634	1.852	1.570	1.325	0.738	3.026	3.023
1887	1.676	1.868	1.616	1.354	0.763	3.121	3.093
1888	1.760	1.900	1.662	1.340	0.723	3.235	3.012
1889	1.758	1.894	1.691	1.272	0.753	3.382	3.134
1890	1.846	1.947	1.725	1.345	0.813	3.369	3.115
1891	1.887	1.994	1.703	1.329	0.768	3.341	3.184
1892	1.863	2.045	1.759	1.246	0.814	3.233	3.424
1893	1.823	2.044	1.827	1.295	0.810	3.203	3.195
1894	1.897	2.118	1.852	1.270	0.899	3.386	3.043
1895	1.874	2.045	1.915	1.281	0.903	3.461	3.347
1896	1.809	2.120	1.954	1.309	0.844	3.571	3.219
1897	1.985	2.072	1.978	1.243	0.853	3.583	3.462
1898	2.042	2.184	2.031	1.345	1.003	3.721	3.471
1899	2.199	2.283	2.073	1.367	0.918	3.838	3.721
1900	2.291	2.323	2.131	1.436	0.948	3.775	3.757
1901	2.447	2.245	2.051	1.519	0.969	3.740	4.100
1902	2.608	2.190	2.068	1.466	0.907	3.803	4.060
1903	2.630	2.253	2.151	1.525	0.958	3.731	4.180
1904	2.592	2.283	2.204	1.527	0.954	3.722	4.050
1905	2.775	2.320	2.219	1.600	0.929	3.799	4.263
1906	3.013	2.334	2.255	1.647	1.042	3.892	4.665
1907	3.025	2.453	2.321	1.818	1.065	3.932	4.652
1908	2.799	2.450	2.328	1.847	1.059	3.739	4.189
1909	2.999	2.515	2.345	1.974	1.045	3.791	4.608
1910	3.179	2.400	2.396	1.884	1.047	3.875	4.559
1911	3.311	2.569	2.447	1.988	1.089	3.958	4.634
1912	3.474	2.772	2.527	1.992	1.112	4.001	4.776

1913	3.514	2.728	2.607	2.072	1.114	4.136	4.868
1914	3.286	2.534	2.194	2.054	1.066	4.141	4.408
1915	3.420	2.543	2.080	2.271	1.134	4.445	4.467
1916	3.713	2.712	2.107	2.536	1.294	4.525	5.013
1917	3.816	2.332	2.119	2.666	1.327	4.556	4.820
1918	3.564	1.876	2.143	2.740	1.331	4.588	5.197
1919	3.282	2.213	1.925	2.298	1.453	4.094	5.217
1920	3.196	2.540	2.080	2.089	1.362	3.838	5.099
1921	2.848	2.420	2.289	2.033	1.494	3.498	4.889
1922	3.182	2.842	2.479	2.116	1.471	3.654	5.088
1923	3.332	2.955	2.046	2.221	1.454	3.750	5.661
1924	3.330	3.290	2.379	2.223	1.475	3.877	5.724
1925	3.619	3.280	2.627	2.348	1.514	4.054	5.770
1926	3.760	3.345	2.680	2.353	1.504	3.889	6.064
1927	4.063	3.270	2.931	2.281	1.502	4.188	6.040
1928	4.338	3.488	3.040	2.424	1.604	4.221	6.033
1929	4.246	3.708	3.013	2.487	1.509	4.336	6.336
1930	4.025	3.567	2.955	2.345	1.499	4.287	5.667
1931	3.344	3.334	2.717	2.313	1.600	4.049	5.155
1932	3.055	3.117	2.500	2.371	1.731	4.057	4.436
1933	2.802	3.337	2.644	2.338	1.709	4.158	4.319
1934	3.063	3.300	2.870	2.329	1.732	4.418	4.629
1935	3.278	3.216	3.062	2.533	1.827	4.569	4.965
1936	3.422	3.341	3.310	2.518	1.889	4.755	5.639
1937	3.709	3.532	3.643	2.671	1.984	4.899	5.880
1938	3.771	3.516	3.973	2.669	2.278	4.938	5.568
1939	3.959	3.773	4.248	2.835	2.328	4.934	5.961
1940	4.458	3.182	4.243	2.822	2.328	5.402	6.384
1941	4.973	2.605	4.486	2.763	2.308	5.895	7.437
1942	5.780	2.347	4.510	2.712	2.264	6.019	8.746
1943	5.968	2.251	4.629	2.446	2.267	6.102	10.198
1944	6.116	1.906	4.785	1.982	2.136	5.835	10.900
1945	5.862	2.026	3.492	1.547	1.081	5.560	10.581
1946	5.696	3.035	1.966	2.013	1.169	5.315	8.466
1947	5.824	3.257	2.248	2.351	1.248	5.204	8.066
1948	5.801	3.458	2.595	2.466	1.397	5.315	8.232
1949	5.807	3.894	2.964	2.629	1.457	5.481	8.097
1950	6.113	4.159	3.338	2.840	1.563	5.623	8.610
1951	6.278	4.379	3.652	3.034	1.732	5.812	9.326
1952	6.550	4.460	3.937	3.241	1.905	5.785	9.519
1953	6.672	4.552	4.223	3.454	2.017	6.039	9.737
1954	6.454	4.735	4.505	3.611	2.101	6.254	9.443
1955	6.878	4.968	5.002	3.799	2.276	6.447	9.802
1956	7.234	5.166	5.299	3.938	2.396	6.513	9.821
1957	7.249	5.422	5.533	4.136	2.548	6.605	9.809
1958	7.190	5.375	5.666	4.322	2.671	6.581	9.582
1959	7.296	5.602	6.017	4.554	2.888	6.807	9.974
1960	7.339	5.938	6.479	4.764	3.237	7.183	9.995
1961	7.422	6.199	6.691	5.132	3.594	7.363	10.101
1962	7.803	6.495	6.925	5.497	3.879	7.376	10.475

1963	8.057	6.727	7.048	5.848	4.165	7.615	10.768
1964	8.435	7.091	7.444	6.029	4.603	7.976	11.241
1965	8.833	7.363	7.765	6.119	4.818	8.127	11.758
1966	9.260	7.683	7.921	6.396	5.283	8.236	12.218
1967	9.361	7.979	7.894	6.807	5.814	8.380	12.360
1968	9.697	8.258	8.303	7.333	6.541	8.686	12.749
1969	10.084	8.766	8.837	7.703	7.206	8.831	12.987
1970	10.202	9.185	9.203	7.803	7.889	8.987	12.825
1971	10.656	9.534	9.369	7.890	8.122	9.117	13.074
1972	11.143	9.869	9.704	8.056	8.681	9.405	13.589
1973	11.866	10.323	10.110	8.573	9.287	10.058	14.103
1974	12.225	10.574	10.124	8.977	9.001	9.885	13.875
1975	12.358	10.498	10.000	8.688	9.088	9.808	13.599
1976	12.949	10.899	10.594	9.214	9.453	10.081	14.126
1977	13.251	11.200	10.929	9.485	9.857	10.322	14.606
1978	13.714	11.526	11.256	9.799	10.267	10.699	15.192
1979	14.110	11.848	11.716	10.355	10.709	10.986	15.322
1980	14.143	11.980	11.837	10.773	11.098	10.723	15.130
1981	14.484	12.054	11.835	10.860	11.441	10.581	15.310
1982	13.880	12.293	11.766	10.870	11.686	10.772	14.765
1983	14.206	12.322	11.984	10.956	11.972	11.147	15.194
1984	14.995	12.435	12.371	11.251	12.493	11.360	16.131
1985	15.588	12.617	12.641	11.520	13.003	11.747	16.581
1986	15.947	12.856	12.927	11.792	13.242	12.123	16.884
1987	16.484	13.049	13.145	12.121	13.733	12.659	17.339
1988	17.095	13.447	13.569	12.574	14.456	13.150	17.966
1989	17.596	13.769	13.989	12.951	15.101	13.460	18.357

Source: raw data from Maddison (1979, 1991) and Summers and Heston (Penn World Tables, Mark 6, NBER gopher). Computations by the author. See the main text for details.

Table 8.17

GDP per head in nine industrialised countries (1870-1989)

(1985 US 1000 relative dollars)

	Australia	Austria	Belgium	Denmark	Finland	Netherlands	Norway	Sweden	Switzerland
1870	3.123	1.433	2.104	1.547	0.933	2.064	1.190	1.316	1.848
1871	2.903	1.523	2.093	1.540	0.922		1.201	1.332	
1872	3.128	1.521	2.203	1.613	0.937		1.269	1.381	
1873	3.365	1.472	2.200	1.589	0.976		1.287	1.492	
1874	3.115	1.524	2.255	1.619	0.985		1.318	1.533	
1875	3.538	1.518	2.232	1.631	0.991		1.343	1.452	
1876	3.410	1.539	2.243	1.646	1.030		1.363	1.531	
1877	3.437	1.578	2.250	1.581	0.991		1.351	1.464	
1878	3.650	1.616	2.295	1.624	0.959		1.292	1.439	
1879	3.470	1.591	2.298	1.660	0.955		1.292	1.400	
1880	3.529	1.600	2.391	1.686	0.945	2.306	1.320	1.461	
1881	3.636	1.651	2.395	1.687	0.908		1.328	1.465	
1882	3.285	1.647	2.447	1.731	0.984		1.324	1.523	
1883	3.637	1.700	2.453	1.777	1.006		1.320	1.533	
1884	3.532	1.730	2.447	1.766	0.998		1.339	1.563	
1885	3.638	1.705	2.448	1.757	1.007		1.345	1.544	
1886	3.563	1.745	2.460	1.805	1.044		1.343	1.526	
1887	3.825	1.850	2.535	1.850	1.044		1.350	1.524	
1888	3.741	1.831	2.533	1.846	1.065		1.405	1.559	
1889	3.951	1.798	2.636	1.854	1.085		1.448	1.634	
1890	3.723	1.880	2.674	1.949	1.130	2.569	1.477	1.651	2.339
1891	3.938	1.928	2.649	1.974	1.105		1.478	1.666	
1892	3.391	1.951	2.685	2.007	1.063		1.501	1.697	
1893	3.145	1.943	2.696	2.031	1.097		1.532	1.696	
1894	3.194	2.035	2.706	2.053	1.145		1.523	1.718	
1895	3.069	2.068	2.740	2.141	1.221		1.519	1.786	
1896	3.244	2.078	2.771	2.191	1.284		1.544	1.873	
1897	3.010	2.101	2.798	2.212	1.329		1.598	1.922	
1898	3.425	2.197	2.820	2.217	1.365		1.579	1.944	
1899	3.379	2.221	2.852	2.280	1.315		1.601	1.971	2.540
1900	3.532	2.217	2.911	2.331	1.365	2.842	1.610	2.026	2.591
1901	3.379	2.204	2.902	2.399	1.339	2.735	1.635	1.990	2.637
1902	3.365	2.266	2.917	2.427	1.302	2.828	1.654	1.975	2.678
1903	3.588	2.263	2.943	2.543	1.379	2.861	1.638	2.112	2.720
1904	3.780	2.274	2.981	2.569	1.416	2.793	1.630	2.120	2.761
1905	3.769	2.378	3.028	2.585	1.425	2.844	1.641	2.129	2.801
1906	3.967	2.444	3.056	2.628	1.468	2.939	1.693	2.251	2.840
1907	4.063	2.569	3.068	2.694	1.501	2.909	1.749	2.283	2.875
1908	4.131	2.554	3.068	2.744	1.496	2.867	1.793	2.258	2.913
1909	4.384	2.521	3.098	2.814	1.542	2.967	1.824	2.222	2.949
1910	4.586	2.532	3.171	2.863	1.560	2.964	1.875	2.358	2.985
1911	4.492	2.589	3.237	2.980	1.586	3.008	1.919	2.375	3.019

1912	4.440	2.697	3.282	2.946	1.655	3.131	1.988	2.424	3.053
1913	4.523	2.667	3.292	3.022	1.727	3.178	2.078	2.450	3.086
1914	4.080	2.214	3.062	3.176	1.637	3.061	2.103	2.411	3.063
1915	3.812	2.041	3.009	2.919	1.539	3.081	2.170	2.396	3.074
1916	4.093	2.023	3.183	3.007	1.549	3.105	2.192	2.348	3.092
1917	4.190	1.989	2.746	2.795	1.293	2.847	1.994	2.044	2.751
1918	4.143	1.966	2.234	2.672	1.128	2.631	1.900	2.004	2.747
1919	3.977	1.737	2.644	2.982	1.357	3.235	2.200	2.112	2.937
1920	4.147	1.856	3.091	3.084	1.510	3.312	2.310	2.217	3.121
1921	4.468	2.039	3.165	2.947	1.542	3.478	2.093	2.116	3.045
1922	4.409	2.214	3.443	3.210	1.684	3.610	2.314	2.299	3.339
1923	4.402	2.187	3.537	3.512	1.789	3.638	2.349	2.411	3.525
1924	4.639	2.434	3.618	3.488	1.820	3.843	2.324	2.477	3.645
1925	4.693	2.591	3.611	3.373	1.905	3.949	2.451	2.558	3.897
1926	4.542	2.626	3.702	3.542	1.957	4.205	2.491	2.693	4.069
1927	4.440	2.697	3.811	3.588	2.092	4.320	2.575	2.769	4.262
1928	4.289	2.814	3.978	3.686	2.215	4.490	2.650	2.893	4.464
1929	4.187	2.846	3.912	3.910	2.223	4.465	2.886	3.061	4.580
1930	3.937	2.759	3.853	4.114	2.181	4.337	3.086	3.115	4.518
1931	3.725	2.530	3.762	4.128	2.111	4.070	2.829	2.993	4.299
1932	3.835	2.262	3.566	3.982	2.086	3.952	2.999	2.901	4.130
1933	4.022	2.180	3.623	4.076	2.210	3.890	3.054	2.945	4.316
1934	4.145	2.194	3.579	4.161	2.445	3.772	3.134	3.158	4.306
1935	4.282	2.237	3.788	4.221	2.530	3.869	3.252	3.349	4.273
1936	4.453	2.304	3.803	4.295	2.682	4.074	3.433	3.533	4.274
1937	4.667	2.428	3.840	4.367	2.815	4.264	3.537	3.690	4.464
1938	4.659	2.739	3.740	4.439	2.937	4.121	3.605	3.739	4.622
1939	4.627	3.152	3.986	4.617	2.788	4.352	3.753	3.979	4.601
1940	4.880	3.047	3.531	3.941	2.635	3.792	3.398	3.843	4.627
1941	5.373	3.245	3.373	3.523	2.718	3.557	3.460	3.888	4.566
1942	5.940	3.065	3.093	3.566	2.722	3.224	3.304	4.098	4.418
1943	6.097	3.128	3.022	3.913	3.024	3.124	3.214	4.241	4.341
1944	5.827	3.195	3.183	4.270	3.014	2.079	3.018	4.339	4.404
1945	5.469	1.327	3.354	3.903	2.822	2.109	3.348	4.405	5.607
1946	5.220	1.505	3.540	4.450	3.013	3.498	3.664	4.828	5.917
1947	5.275	1.667	3.715	4.649	3.041	3.962	4.035	4.886	6.547
1948	5.515	2.127	3.888	4.752	3.237	4.309	4.307	4.979	6.594
1949	5.728	2.534	4.020	4.920	3.390	4.616	4.347	5.107	6.848
1950	5.900	2.852	4.228	5.224	3.480	4.706	4.541	5.332	6.556
1951	6.004	3.046	4.445	5.147	3.734	4.745	4.705	5.506	7.002
1952	5.915	3.047	4.387	5.202	3.817	4.781	4.825	5.544	6.974
1953	5.959	3.179	4.505	5.459	3.802	5.134	5.000	5.659	7.135
1954	6.185	3.502	4.665	5.605	4.089	5.417	5.197	5.858	7.450
1955	6.379	3.888	4.862	5.543	4.252	5.750	5.248	5.986	7.860
1956	6.449	4.144	4.971	5.618	4.340	5.882	5.473	6.177	8.256
1957	6.432	4.385	5.031	5.825	4.503	5.967	5.582	6.416	8.455
1958	6.373	4.532	4.994	5.952	4.487	5.872	5.482	6.409	8.155
1959	6.852	4.647	5.119	6.320	4.712	6.079	5.715	6.566	8.542
1960	7.017	5.016	5.381	6.653	5.097	6.504	5.991	6.874	9.010
1961	7.035	5.253	5.621	7.032	5.448	6.437	6.316	7.229	9.610

1962	7.140	5.348	5.880	7.371	5.573	6.781	6.442	7.595	9.818
1963	7.488	5.533	6.092	7.361	5.715	6.933	6.634	7.934	10.082
1964	7.862	5.830	6.457	7.981	5.981	7.407	6.916	8.610	10.438
1965	8.060	5.964	6.630	8.278	6.276	7.689	7.224	8.998	10.646
1966	8.225	6.255	6.796	8.434	6.401	7.797	7.439	9.172	10.793
1967	8.541	6.418	7.023	8.648	6.506	8.116	7.837	9.356	10.986
1968	8.921	6.682	7.294	8.940	6.624	8.548	7.943	9.761	11.237
1969	9.361	7.080	7.761	9.459	7.264	8.992	8.231	9.937	11.738
1970	9.693	7.543	8.232	9.575	7.837	9.388	8.339	10.491	12.227
1971	9.878	7.890	8.502	9.763	7.984	9.669	8.663	10.818	12.609
1972	10.163	8.326	8.917	10.220	8.549	9.889	9.040	10.764	12.890
1973	10.428	8.682	9.417	10.528	9.073	10.267	9.346	11.293	13.187
1974	10.464	9.013	9.773	10.382	9.297	10.590	9.773	11.620	13.355
1975	10.554	8.989	9.600	10.282	9.364	10.488	10.124	11.869	12.370
1976	10.716	9.405	10.118	10.920	9.360	10.937	10.762	11.953	12.395
1977	10.797	9.807	10.155	11.064	9.346	11.124	11.100	11.719	12.735
1978	11.008	9.865	10.424	11.192	9.520	11.328	11.559	11.890	12.768
1979	11.309	10.335	10.638	11.209	10.186	11.517	12.103	12.321	13.057
1980	11.500	10.641	11.085	11.496	10.696	11.524	12.556	12.577	13.727
1981	11.610	10.608	10.965	11.395	10.819	11.366	12.639	12.567	13.848
1982	11.441	10.712	11.129	11.748	11.144	11.154	12.635	12.698	13.613
1983	11.514	10.977	11.172	12.054	11.407	11.269	13.162	12.919	13.647
1984	11.998	11.116	11.414	12.588	11.694	11.578	13.896	13.423	13.835
1985	12.411	11.386	11.507	13.123	12.034	11.827	14.583	13.701	14.339
1986	12.638	11.500	11.688	13.507	12.246	11.999	15.135	13.979	14.667
1987	12.900	11.711	11.897	13.395	12.696	12.048	15.577	14.339	14.869
1988	13.180	12.162	12.366	13.333	13.306	12.298	15.665	14.609	15.292
1989	13.587	12.585	12.876	13.512	13.934	12.738	16.503	14.809	15.577

Source: raw data from Maddison (1979, 1991) and Summers and Heston (Penn World Tables, Mark 6, NBER gopher). Computations by the author. See the main text for details.

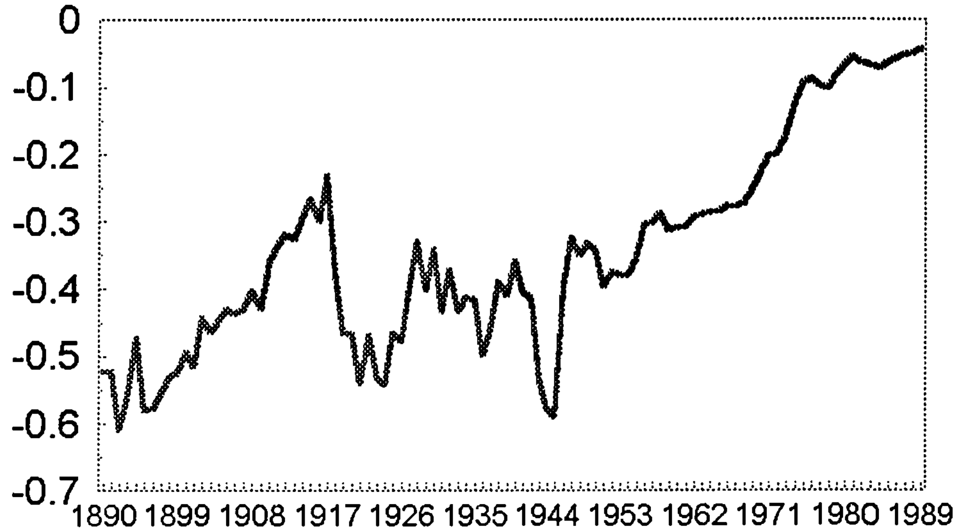

Figure 8.1
Difference between Canadian and US GDP per head

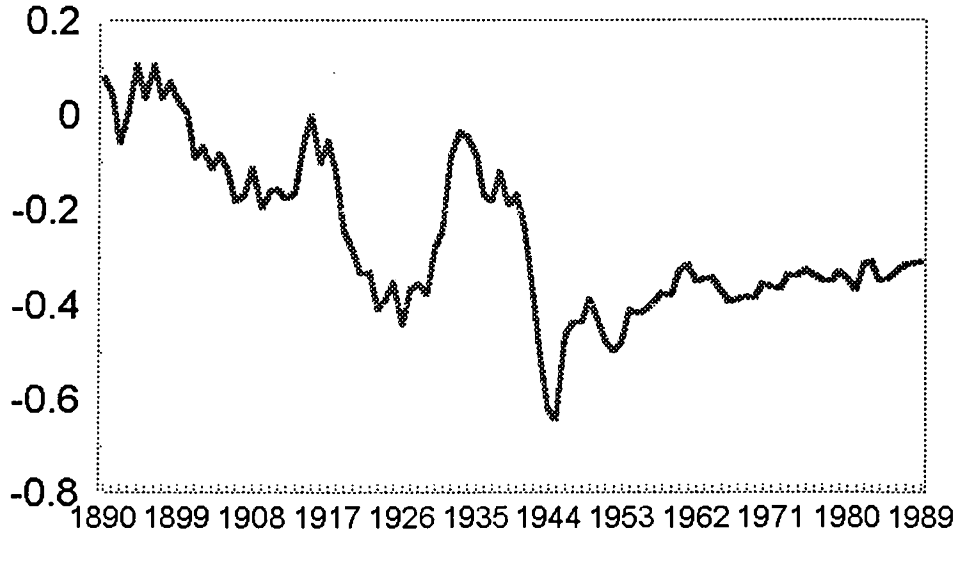

Figure 8.2
Difference between British and US GDP per head

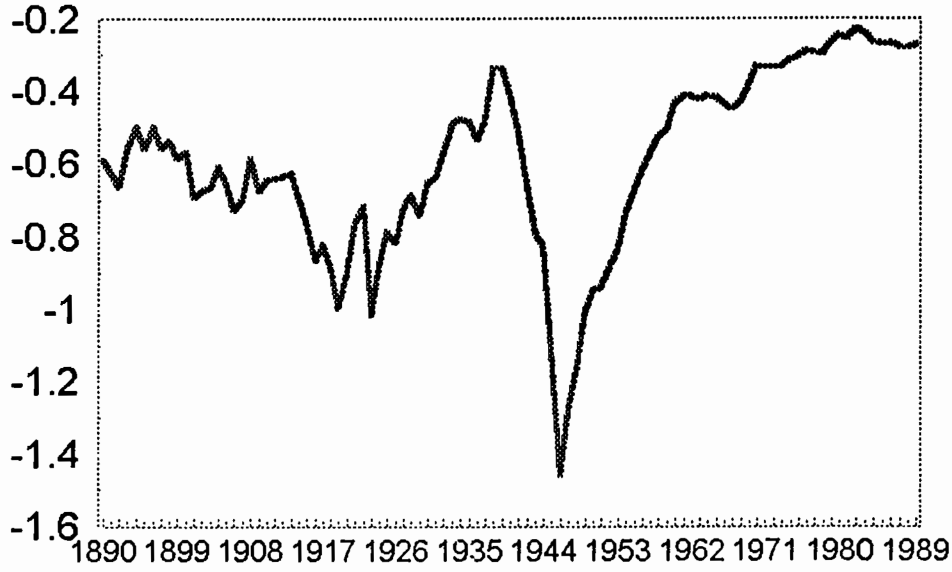

Figure 8.3
Difference between German and US GDP per head

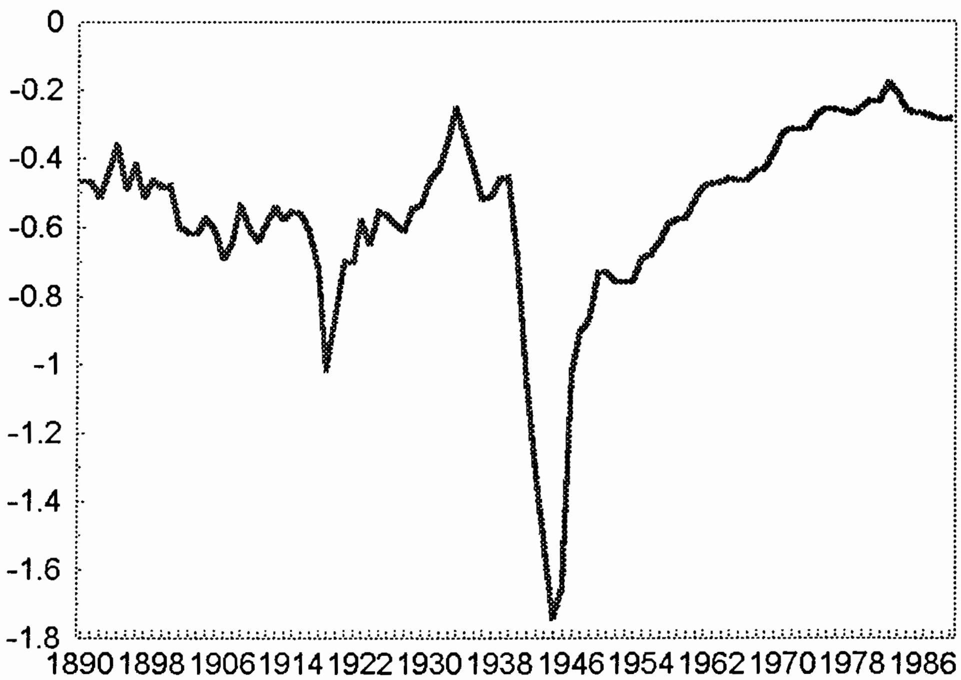

Figure 8.4
Difference between French and US GDP per head

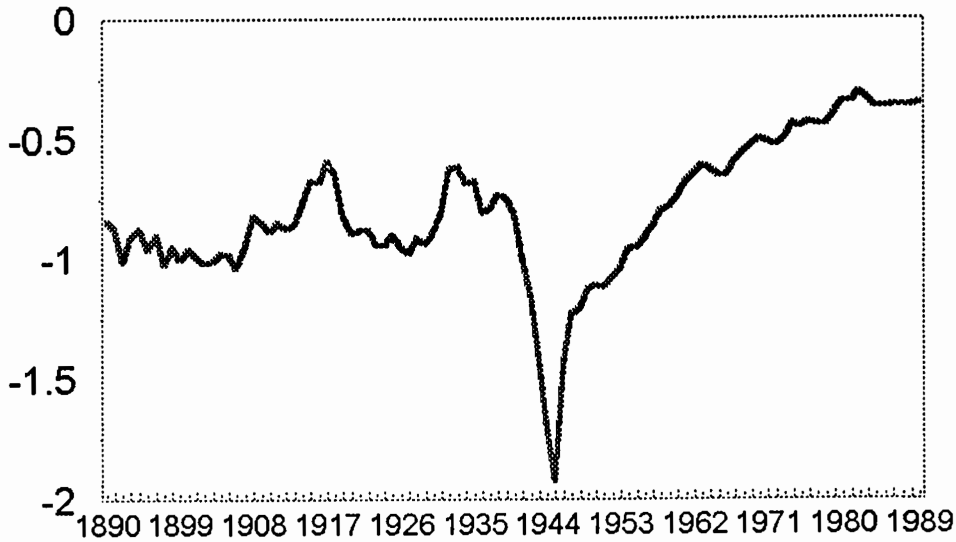

Figure 8.5
Difference between Italian and US GDP per head

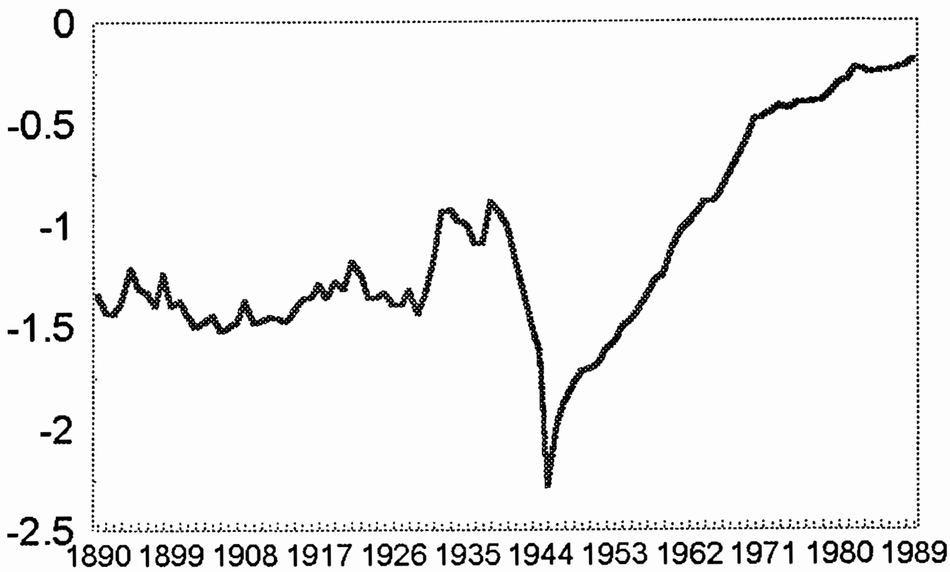

Figure 8.6
Difference between Japanese and US GDP per head

Conclusion

1. The starting point

The importance of convergence in Economics was shown in chapter 1.

In the growth literature, both theoretical and empirical, the controversy on convergence is all too apparent. Different models of economic growth have different implications in terms of convergence. The standard neoclassical exogenous growth model implies a conditional convergence result for incomes per head across countries. Alternative models, including a legion of endogenous growth models developed in the last ten years, usually challenge this result. They suggest instead that growth rates are endogenously determined by different conditions across countries. Consequently, there is no *a priori* reason to expect convergence of any kind; endogenous growth models allow for permanent differences in growth rates, so that the gap between two countries can increase over time.

Technological catch up and diffusion models provide a possible conciliation of these different results. Growth is endogenous in leader countries and exogenous for the followers, in the sense that policies and other variables will only determine their long term relative position towards the leader (but not their long term growth rate.)

International trade and the macroeconomics of EMU (Economic and Monetary Union) are other fields of Economics where the term "convergence" is important:

- the factor prices equalisation theorem states that, under a set of assumptions, factor prices should equalise across countries that trade with each other.
- The Treaty on European Union (also known as the Maastricht Treaty) states that a country should obey four convergence criteria to participate in the EMU. Namely, inflation rates, government deficits, exchange rates and interest rates have to converge according to specific rules and definitions.

All these developments led to a surge in empirical methods and results in measuring convergence of economic series. Particularly, there was a great development in this decade of the empirical testing of the convergence of incomes hypothesis.

One striking feature of this literature is that different empirical methods lead to (apparently) contradictory results when applied to the same data set.

The following stylised facts emerge from recent literature on the empirics of economic growth and convergence:

a) initial value regressions tend to reject the "no unconditional convergence hypothesis" when similar economies are considered and to reject the "no conditional convergence hypothesis" when a large cross-section is used. In other words, when a restricted group of not very different economies are considered (e. g. regions of a country, American states, OECD countries), it is found that there is a statistically significant negative correlation between growth rates and initial incomes. The estimated speed of convergence is usually small and not very different from 0.02 per year.

b) In some cases (e. g. European regions) Markov chains methods do not confirm an unconditional convergence result that is apparent from initial value regressions. Even if there is an empirical correlation between initial conditions and growth rates, without any conditioning variables, the long-run distribution derived from the estimated transition matrices may not show evidence of convergence in the sense that incomes do not tend to pile up in the middle quantiles.

c) Time series techniques (e. g. Dickey-Fuller tests) tend not to reject the null of no convergence (conditional or unconditional) even when economies are similar.

2. The aim of the thesis

Considering the starting point, the thesis had the following objectives:

(1) To evaluate the power and properties of the different tests for convergence under different convergence situations likely to happen in practice.

(2) To understand how and why different tests for convergence applied to the same data can provide results that are in apparent contradiction.

(3) To provide an empirical application that takes (1) and (2) into account and contributes positively to the ever growing literature on the empirics of GDP per head convergence.

Objectives (1) and (2) were pursued using artificial data and Monte Carlo methods. The different convergence situations that were considered were:

- complete unconditional convergence (Chapter 3);
- complete conditional convergence (Chapter 4);
- convergence clubs (Chapter 5);
- limited convergence (Chapter 6);
- time-varying convergence (Chapter 7.)

An empirical application where convergence of incomes per head is assessed for 16 industrialised countries from 1890 to 1989 constituted chapter 8.

3. Main results and conclusions

For ease of exposition, results and conclusions from chapters 3 to 7 are presented separately from the empirical results derived from chapter 8.

3.1 Evaluating methods for measuring convergence: results and conclusions

It is useful to divide methods of testing for convergence between methods that use more than two series at a time ("cross-sectional methods") and methods that use two series at a time only ("time series methods.")

The following main points result from the studies contained in Chapters 3 to 7:

(1) Very different time paths for cross-section dispersion measures are compatible with a situation where all series converge, either conditionally or unconditionally. Even if extreme cases of unconditional convergence (when the series start far from each other) usually imply a declining pattern for the cross-section dispersion measures time path, it must be remembered that a similar pattern can be the result of limited convergence. Also, convergence can occur even if cross-section dispersion increases through time. The information content of this type

of measures is therefore limited.

(2) Cross-sectional methods are likely to provide misleading information or have low power in the presence of "cross-section variation." "Cross-section variation" is a situation where not all the series converge to the same leader series. Two different types of cross-section variation were considered.

If part of the series converges to one leader and the other part converges to a different leader, it is said that there are two different "convergence clubs." If there is only one leader, but only a part of the series converges to it, this is called "limited convergence."

Initial value and random field regressions are not adequate in dealing with this kind of variation:

- on the one hand, the existence of different clubs or limited convergence can well imply a negative cross-section correlation between initial income and growth rates that is mirrored in a significant negative coefficient on an initial value regression. Similarly, the same negative correlation is possibly reflected in a coefficient smaller than one in a random field regression. These possible results should be interpreted with care; the dismissal of the no convergence hypothesis does not imply that all series are converging to the same leader;

- on the other hand, it is also possible that the series are converging in club or limited form, in such a way as not to induce any dismissal of the no convergence hypothesis when the researcher is using an initial value or random field regression tests. Here, it is important guard oneself against a hasty conclusion that convergence is not occurring in any form.

Extensions of initial value regressions, namely Quandt-type tests, have the advantage of directly addressing the possible existence of different clubs. In some occasions they are useful in detecting the existence of a cross-sectional structural break. Nevertheless, they rely on the correlation between the composition of clubs and a set of controlling variables. If this correlation is not very strong, clubs are nor correctly identified.

Markov chain methods also suffer from some important drawbacks in the presence of convergence clubs or limited convergence. When the composition of clubs is highly related to initial income, or when only the richest countries converge, the estimated transition matrices or long-run distribution vectors usually reflect this process more or less accurately. Unfortunately, this is not the case when convergence is not clearly related to initial wealth. In this last case, a convergence process that is limited or implies different clubs can show up as a complete convergence process in the estimated matrices.

(3) Cross-sectional methods are likely to provide misleading information or have low power when convergence is conditional.

Under conditional convergence, the difference between two converging series tends to a constant in the long-run, not necessarily zero. Random field and Markov chains methods were designed to cope with unconditional convergence. It comes as no surprise that they do not detect conditional convergence.

Initial value regressions can include some conditioning variables to allow for conditional convergence. These variables are proxies for the different steady-state levels, which are unobservable. If the proxies are not highly correlated to the steady-state levels the power of the test declines and the speed of convergence is underestimated.

(4) Time series methods (Dickey-Fuller and Kalman filter) are, by construction, robust to "cross-section variation."

The Dickey-Fuller tests and the Kalman filter method use two series at a time. This means that belonging of a series to a club can be directly tested. If series A belongs to the same club as series B, the researcher can test for convergence between series A and B. Since convergence is transitive, it does not matter if one of the series is attracting the other. They can be both attracted by a third series.

(5) The power of time series tests increases when the speed of convergence is higher and when the initial distance between the series is greater, relative to its steady-state level. It also depends positively on the number of observations.

Time series methods are both adequate to test for conditional and unconditional convergence. It does not matter if the difference between any two series tends to zero or to a constant that is different from zero. In this respect, the important variable that is relevant for the power of the test is the gap between the initial and the long run distance.

As usual, time series methods have more power when long series are considered. Monte Carlo studies and the empirical work included in this thesis suggest that they retain their usefulness when forty years are considered in a convergence of income per head study.

(6) Cointegration methods (Dickey-Fuller methods) lose their power dramatically when the convergence process does not start in the first time period. In this case, the Kalman filter method is more successful in detecting convergence.

Dickey-Fuller methods are at their best when the speed of convergence is constant from the first period to the last one. In a probably more realistic setting, convergence starts to occur somewhere after the first period, so there is a structural break. The power of both time series methods declines when the break point is farther from the initial moment, but this is stronger in the Dickey-Fuller test: the Kalman filter method was more robust to structural change.

(7) Cross-sectional methods provide useful information when time is the only source of structural variation. Estimates for the speed of convergence provided by random field or initial value regressions are biased, though.

In the particular case where the speed of convergence is time-varying but equal across all series, convergence being unconditional and complete, initial value and random field regressions and Markov chains methods retain some of their desirable properties. In the instance where convergence starts to take place some periods after the initial period, random field and initial value regressions underestimate the speed of convergence.

3.2 Convergence across industrialised countries (1890-1989): the main findings

Chapter 8 is an empirical investigation on the convergence of GDPs per head across sixteen industrialised countries using long series (a century of annual data.) Dickey-Fuller and Kalman filter tests for convergence were used and the United States were chosen as a benchmark.

The main findings can be summarized by the following points:

(1) The "no convergence hypothesis" was rejected for all countries when the Kalman filter test was used and the whole period was considered. The same hypothesis could not be dismissed for seven countries when using the Dickey-Fuller test. One possible explanation is that structural change occurred during the long period considered.

(2) When two different periods are considered (one before the Second World War and one after it) convergence is found to be more prevalent in the more recent one. This strengthens the structural change explanation mentioned in (1), suggesting that, for a typical country, convergence started to take place between 1890 and 1939, was disrupted by the Second World War and resumed soon after it.

(3) Estimates for the steady-state levels both before and after the war led to the conclusion that the last ones were generally higher. Note, however, that estimated steady-states for France, Germany, Italy and the United Kingdom GDP per head correspond to roughly three quarters of the US level. The fact that almost all the countries were far from this steady-state immediately after the Second World War is probably responsible for the high power of time series methods in this period.

(4) Estimates for the speed of convergence were very different across countries. After the war, they varied from 0.03 for Japan and 0.319 for the United Kingdom (this last country was converging faster to a lower steady-state.) Also, the speed of convergence was generally different when the same country was considered in the two different periods.

Appendix

Some Gauss Routines Used in the Experiments Presented in this Thesis

1. The data generation process

The two following routines are an example of a data generation process programmed in Gauss. This particular routine were used for some of experiments presented in chapter 6 ("Limited Convergence.")

```
/***********************************/
INIT:
/* initialises some variables *********/


t=30;
betac=0.06;

beta=zeros(1,99);

print "proportion of countries to converge";
pcon=con(1,1);

bar=zeros(repno,2);
qu=zeros(repno, 3);
return;

DATAGER:
/* data generation routine *****************************************/
/* limited unconditional convergence */
/* club uniformily distributed */


n=99;
```

```
/* X stores the attracted series */
X=zeros(t,n);

/* Y stores the attractor series */
Y=zeros(t,1);

/*g is the attractor series growth rate */
g=0.02172128;

/*vare is the total variance of shocks */
vare=0.00071070513;

/*initial values */
load ly6089;

X[1,.]=ly6089[1,2:n+1];
Y[1]=ly6089[1,1];

step=round(1/pcon);

i=0;
do while i < 99-step;
        i=i+step;
        beta[i]=betac;
endo;

gx=g; /* growth rates for attracted series */

i=1;
etlx=sqrt(1/3*vare)*rndns(1,n,seed);
etly=sqrt(1/3*vare)*rndns(1,1,seed);
do while i<t;
        i=i+1;
```

```
        ey=sqrt(1/3*vare)*rndns(1,1,seed);
        etx=sqrt(1/3*vare)*rndns(1,n,seed);
        ety=sqrt(1/3*vare)*rndns(1,1, seed);
        ex=sqrt(1/3*vare)*rndns(1,n, seed);
        X[i,.]=X[i-1,.]+gx+beta.*(Y[i-1]-X[i-1,.])+ex+etx-etlx;
        Y[i,1]=Y[i-1]+g+ey+ety-etly;
        etly=ety;
        etlx=etx;
endo;
MXY=Y~X;
return;
/* end of generation routine */
/*****************************************/
```

2. Augmented Dickey-Fuller tests

The two following routines were written to compute Dickey-Fuller or Augmented Dickey Fuller tests for the stationarity of the differences of two series.

```
/**************************************/
DICKEY:

__output=0;
_olsres=1;

if nlags eq 0;
/* DF calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[2:t]-d[1:t-1], d[1:t-1]);
tstat= b[2]/std[2];
endif;

if nlags eq 1;
```

```
/* ADF (1) calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[3:t]-d[2:t-1], d[2:t-1]~(d[2:t-1]-d[1:t-2]));
tstat= b[2]/std[2];
endif;

if nlags eq 2;
/* ADF (2) calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[4:t]-d[3:t-1], d[3:t-1]~(d[3:t-1]-d[2:t-2])~(d[2:t-2]-d[1:t-3]));
tstat= b[2]/std[2];
endif;

if nlags eq 3;
/* ADF (3) calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[5:t]-d[4:t-1], d[4:t-1]~(d[4:t-1]-d[3:t-2])~(d[3:t-2]-d[2:t-3])~
(d[2:t-3]-d[1:t-4]));
tstat= b[2]/std[2];
endif;

if nlags eq 4;
/* ADF (4) calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[6:t]-d[5:t-1], d[5:t-1]~(d[5:t-1]-d[4:t-2])~
(d[4:t-2]-d[3:t-3])~(d[3:t-3]-d[2:t-4])~(d[2:t-4]-d[1:t-5]));
endif;

if nlags eq 5;
/* ADF (5) calculations */
{v, m, b, stb, v, std, s, c, r, res, dw}=
OLS(0, d[7:t]-d[6:t-1], d[6:t-1]~(d[6:t-1]-d[5:t-2])~
(d[5:t-2]-d[4:t-3])~(d[4:t-3]-d[3:t-4])~(d[3:t-4]-d[2:t-5]
```

```
~d[2:t-5]-d[1:t-6]));
endif;

tstat= b[2]/std[2];

n=rows(res);
/*Godfrey's test of residual serial correlation (Microfit Manual, p. 184-5) */

W1 = 0|res[1:n-1];
W2 = W1~(0|0|res[1:n-2]);
W3 = W2~(0|0|0|res[1:n-3]);

if nlags eq 0;
/* DF calculations */
Xd = ones(n,1)~d[1:t-1];
endif;

if nlags eq 1;
/* ADF(1) calculations */
Xd = ones(n,1)~d[2:t-1]~(d[2:t-1]-d[1:t-2]);
endif;

if nlags eq 2;
/* ADF(2) calculations */
Xd = ones(n,1)~d[3:t-1]~(d[3:t-1]-d[2:t-2])~(d[2:t-2]-d[1:t-3]);
endif;

if nlags eq 3;
/* ADF(3) calculations */
Xd = ones(n,1)~d[4:t-1]~(d[4:t-1]-d[3:t-2])~
(d[3:t-2]-d[2:t-3])~(d[2:t-3]-d[1:t-4]);
endif;
```

```
if nlags eq 4;
/* ADF(4) calculations */
Xd = ones(n,1)~d[5:t-1]~(d[5:t-1]-d[4:t-2])~
(d[4:t-2]-d[3:t-3])~(d[3:t-3]-d[2:t-4])~(d[2:t-4]-d[1:t-5]);
endif;

if nlags eq 5;
/* ADF(5) calculations */
Xd = ones(n,1)~d[6:t-1]~(d[6:t-1]-d[5:t-2])~
(d[5:t-2]-d[4:t-3])~(d[4:t-3]-d[3:t-4])~(d[3:t-4]-d[2:t-5]
~d[2:t-5]-d[1:t-6]);
endif;

M = eye(n)-Xd*inv(Xd'Xd)*Xd';

q1 = (n)*(res'*W1*inv(W1'*M*W1)*W1'*res)/(res'*res);
q2 = (n)*(res'*W2*inv(W2'*M*W2)*W2'*res)/(res'*res);
q3 = (n)*(res'*W3*inv(W3'*M*W3)*W3'*res)/(res'*res);

q1s=cdfchic(q1,1);
q2s=cdfchic(q2,2);
q3s=cdfchic(q3,3);

return;

PRINTI:

df8[rep,.]=tstat~q1s~q2s~q3s;
return;
```

3. Initial value regressions

The routine reproduced below was used to perform unconditional initial value regressions.

```
/***************************/
BARRO:

/* initial value regression routine */

gr=(-(Y[1]~X[1,.])'+(Y[t]~X[t,.])')/(t-1);

__output=0;
__con=1;
{vnam, m, bb, stb, vc, stderr, sigma, cx, rsq, res, dw} =
OLS(0, gr, Y[1]|X[1,.]');

k1=(t-1)*bb[2]+1;

format 8,3;

tbarro=(bb[2])/stderr[2];

if k1 > 0;
      lambda=1-k1^(1/(t-1));
else;
      lambda=-100;

endif;

bar[rep,.]=lambda~tbarro;
return;
/* end of initial value regression routine */
```

4. Random field regressions

This routine was employed in computing random fields regressions using a set of series.

```
/***********************************/
QUAH:
/* random fields routine */

format /mb1 /ros 8,3;

Xb=MXY-MXY[.,1];

X=VECR(Xb[.,2:n]);
clear Xb;

__con=0;
__output=0;

{ vnam,m,b,stb,vc,sterr,sigma,cx,rsq,resid,dwstat }=
OLS(0,X[n:(n-1)*t,1],X[1:(n-1)*(t-1),1]);

/* saving estimated coeff, t-stat and unit root stat */

qu[rep,.]=b[1,1]~b[1,1]/sterr[1,1]~sqrt(n/2)*T*(b-1);

return;
/* end of random fields routine **********************************/
/*********************************************************/
```

5. The Kalman filter

The Kalman filter method was implemented in Gauss with the following code.

```
/* Kalman filter estimation routine ***************************************/

HALL:
```

```
errf=0;
logl=0;
varm=0;

/*** initial values for the parameter **********/
Bi=1;

/*** initial value for the disturbance variance ***/
Qi=1000;
/*** initial value for the decaying parameter ***/
phi=1000;


Pin=100000; /* setting uncertainty on the initial parameter value*/

hpari=Qi|phi; /* initial value for log-likelihood function */

param=zeros(t,1);

_opgtol=0.0001;
_opstmth="steep golden";
_opmdmth="newton stepbt";
_opmiter=500;
_grnum=20;
_grsca=0.4;
_grstp=0.5;
__output=0;
{hparf,logl,g,retcode}=optmum(&L,real(hpari));
hparfa=abs(hparf);
H=hessp(&L,hparfa);
Cov=inv(H);
@testing for minimum@
ev=eig(Cov);
```

```
if (ev GT 0);

        h00210[itern]=(hparfa[2]-1000)/sqrt(cov[2,2]);
        save h00210;
else;
        itern=itern-1;
endif;

return;

/* end of Kalman filter estimation routine ****************************/
/**********************************************************************/

PROC(1)=L(hpar);

local Pp,j,Dt,X1t,X2t,Q,Pf,P,f,v,B,slogf,sv2f, phia, Qp;

errf=0;
B=zeros(t,1);
B[1,.]=Bi';
Pp=Pin;
phia=(hpar[2,1]/1000)^2;
Qp=(hpar[1,1]/1000)^2/phia;

Dt=MXY[.,1]-MXY[.,2];
j=1;
logl=0;
slogf=0;
sv2f=0;
do while j<t;
      Q=phia*Qp;
      j=j+1;
      Pf=Pp+Q;
```

```
        f=Pf+1;
            if f<=0;
                errf=f;
                closeall;
                logl=(rndu(1,1)+1)*100000;
                retp(logl);
            endif;
        P=Pf-(Pf^2)/f;
        v=Dt[j,1]-B[j-1,.]';
        B[j,.]=B[j-1,.]+(v*Pf)/f;
        Pp=P;
        Qp=Q;
            if j>1;
                slogf=slogf+ln(f);
                sv2f=sv2f+(v^2)/f;
            endif;
    endo;

    closeall;
    logl=slogf+(t-1)*ln(1/(t-1)*sv2f);
    varm=(t-1)^(-1)*sv2f;
    param=B;
    retp(logl);

    endp;
    /***********************************/
```

6. Markov chains

A routine used to estimate Markov chains is reproduced below.

```
/****************/
/* Markov chains routine **********************/
```

```
MARKOV:
format /mb1 /ros 8,3;
/* g = no. of groups to consider */
g=5;

/* grida =  grid that divides initial values into 5 equal sized groups */
grida= {0.25, 0.5, 1.0, 2.0};

grid=0|grida|100;

/*** full period transition matrix ***/

/*** transform data (data was read simulated in logs) *****/

ir=exp(MXY[1,.]);
fr=exp(MXY[t,.]);
ip=ir/meanc(ir');
fp=fr/meanc(fr');

/*** full period transition matrix estimation ****/


mf=zeros(g,g);

i=0;

do while i < g;

j=0;
i=i+1;
        do while j < g;
                j = j+1;
                v = grid[i,1]|grid[i+1,1];
```

```
            ii = indexcat(ip',v);
            v = grid[j,1]|grid[j+1,1];
            fi = indexcat(fp',v);
            trans = intrsect(ii,fi,1);
            ptrans=rows(trans)/rows(ii);
            mf[i,j]=(1-ismiss(trans))*ptrans;
      endo;
endo;


/* estimating the full period period long run transition matrix ***/

i=0;

mft=mf';
mfl=mft;
do while i < 5000;
      i=i+1;
      mfl=mfl*mft;
endo;


/*** one period transition matrix ****/

/**** one period transition matrix estimation ****************/

m1a=zeros(g,g);

tc=0;
ng=zeros(g,1);
do while tc<t-1;

      tc=tc+1;
```

```
        ir=exp(MXY[tc,.]);
        fr=exp(MXY[tc+1,.]);
        ip=ir/meanc(ir');
        fp=fr/meanc(fr');
        clear ir, fr;
        i=0;

        do while i < g;

        j=0;
        i=i+1;
        v = grid[i,1]|grid[i+1,1];
        ii = indexcat(ip',v);
        if ismiss(ii) ne 1;
              ng[i,1]=ng[i,1]+1;
              do while j < g;
                    j = j+1;
                          v = grid[j,1]|grid[j+1,1];
                          fi = indexcat(fp',v);
                          trans = intrsect(ii,fi,1);
                          ptrans=rows(trans)/rows(ii);
                          m1a[i,j]=m1a[i,j]+(1-ismiss(trans))*ptrans;
              endo;
        endif;
        endo;
endo;


/* one period transition matrix **********************/

m1=m1a./ng;

/* estimating the one period long run transition matrix ***********/
```

```
m1t=m1';
i=0;
m1l=m1t;
do while i < 5000;
      i=i+1;
      m1l=m1l*m1t;
endo;
/*estimating the long run distributions */
/*final distribution*/
fd=zeros(5,1);
i=0;
do while i < 5;
         i=i+1;
         ind=indexcat(fp', grid[i:i+1]);
         fd[i]=(1-ismiss(ind))*(1/(n+1))*rows(ind);
endo;
/* long run disributions */
ld1=m1l*fd;
ldf=mfl*fd;

/* average matrices */


ld1m = ld1m + (1/repno)*ld1;
ldfm = ldfm + (1/repno)*ldf;


m1m=m1m+(1/repno)*m1;
m1lm=m1lm+(1/repno)*m1l;
mfm=mfm+(1/repno)*mf;
mflm=mflm+(1/repno)*mfl;


return;
/* end of Markov chains generation routine ***********************************/
```

References

Abramovitz, M. (1986), "Catching up, forging ahead, and falling behind", Journal of Economic History, 46, 2, June, 385-406.

Abramovitz, M. (1990), "The catch-up factor in postwar economic growth", Economic Inquiry, 28, 1, January, 1-18.

Aghion, P. and P. Howitt (1992), "A model growth through creative destruction", Econometrica, 60, 2, March, 323-351.

Alam, M. S. and A. Naseer (1992), "Convergence and polarization: testing for an inverted-u relation between growth rates and GDP per capita", Applied Economics, 24, 3, March, 363-366.

Anderton, R., R. Barrell, and J. W. in't Veld (1991), "Macroeconomic convergence in Europe", National Institute Economic Review, 138, November, 51-62.

Arrow, K. J. (1962), "The economic implications of learning by doing", Review of Economic Studies, 29, June, 155-73.

Barnett, W., B. Cornett, C. d'Aspremont, J. Gabsewickz and A. Mas-Colell, eds. (1991), Equilibrium Theory and Applications, Proceedings of the Sixth International Symposium in Economic Theory and Econometrics, Cambridge University Press, Cambridge.

Barrell, R., ed., (1992), Economic Convergence and Monetary Union in Europe, Sage, London.

Barro, R. J. (1991), "Economic growth in a cross-section of countries", Quarterly Journal of Economics, 106, 2, May, 407-443.

Barro, R. and X. Sala-i-Martin (1992a), "Convergence", Journal of Political Economy, 100, 2, 223-251.

Barro, R. and X. Sala-i-Martin (1992b), "Convergence across states and regions", in Cukierman, A., Z. Hercowitz, and L. Leiderman (1992), 141-195.

Barro, R. and X. Sala-i-Martin (1992c), "Public finance in models of economic growth", Review of Economic Studies, 59, 4, October, 645-661.

Barro, R. and X. Sala-i-Martin (1995), Economic Growth, McGraw-Hill, New York.

Baumol, W. (1986), "Productivity growth, convergence, and welfare: what the long-run data show", American Economic Revew, 76, 5, September - December, 1072-1085.

Baumol, W and E. Wolf (1988), "Productivity growth, convergence, and welfare: reply", American Economic Review, 78, 5, December, 1155-1159.

Bayoumi, T. (1991), The effect of the ERM on Participating Economies, IMF working paper no. 86.

Bayoumi, T. and B. Eichengreen (1992a), Shocking Aspects of European Monetary Unification, CEPR discussion paper no. 643, London.

Bayoumi, T. and B. Eichengreen (1992b), Is There a Conflict Between EC Enlargement and European Monetary Unification?, CEPR discussion paper no. 646, London.

Bayoumi, T. and G. Sterne (1993), Temporary Cycles or Volatile Trends?: Economic Fluctuations in 21 OECD Countries, Bank of England working paper no. 13, London.

Belmont European Policy Centre (1991), The New Treaty on European Union - Vol. I: A Practical Guide, Belmont European Policy Centre, Brussels.

Ben-David, D. (1994), Convergence Clubs and Diverging Economies, CEPR discussion paper no. 922, London.

Ben-David, D. (1995), Trade and Convergence Among Countries, CEPR discussion paper no. 1126, London.

Bernard, A. B. and S. N. Durlauf (1991), Convergence of International Output Movements, NBER working paper no. 3717, Cambridge, Massachusetts.

Bishop, J., J. Formby and P. D. Thistle (1992), "Convergence and North-South income distributions, 1969-1979", American Economic Review, 82, 1, March, 262-272.

Blanchard, O. and D. Quah (1989), "The dynamic effects of aggregate demand and supply disturbances", American Economic Review, 79, 4, September, 655-673.

Boltho, A. and G. Holtham (1992), "The assessment: new approaches to economic growth", Oxford Review of Economic Policy, 8, 4, Winter, 1-14.

Breiman, L, J. Friedman, R. Olshen and C. Stone (1984), Classification and Regression Trees, Wadworth International Group, Belmont, California.

Breusch, T. S and M. Wickens (1988), "Dynamic specification, the long-run and the estimation of transformed regression models", Economic Journal, 98, 390 (Supplement), 189-205.

Burgman, T. and J. M. Geppert (1993), "Factor price equalisation: a cointegration approach", Weltwirtschaftliches Archiv, 129, 3, 473-487.

Carlino, G. A., (1992), "Are regional per capita earnings diverging?", Business Review, Federal Reserve Bank of Philadelphia, March/April, 3-12.

Carlino, G. and L. Mills (1991), Have Regional Per-capita Incomes Converged?, Federal Reserve Bank of Philadelphia working paper no. 91-18.

Cashin, P. (1993), Economic Growth and Convergence Across the Seven Colonies of Australia: 1961-1991, Economic Growth Center, discussion paper no. 688, Yale University.

Chatterji, M. (1992), "Convergence clubs and endogenous growth", Oxford Review of Economic Policy, 8, 4, Winter, 57-69.

Chiang, A. (1992), Elements of Dynamic Optimization, McGraw-Hill, London.

Cohen, D., (1992), Tests of the "Convergence Hypothesis": A Critical Note, CEPR discussion paper no. 691, London.

Cohen, D. (1993), Two notes on Economic Growth and the Solow Model, CEPR discussion paper no. 780, London.

Costello, D. (1993), "A cross-country, cross-industry comparison of productivity growth", Journal of Political Economy, 101, 2, April, 207-222.

Coughlin, C. and T. Mandelbaum (1989), "Have federal spending and taxation contributed to the divergence of state per capita incomes in the 1980s?", Federal Reserve Bank of St. Louis Review, 71, 4, July/August, 29-42.

Crafts, N. and G. Toniolo (1995), Post-war Growth: An Overview, CEPR discussion paper no. 1095, London.

Craig, C., S. Douglas, A. Grein (1992), "Patterns of convergence and divergence among industrialized nations: 1960-1988", Journal of International Business Studies, 23, 4, 773-787.

Cukierman, A., Z. Hercowitz, and L. Leiderman (1992), Political Economy, Growth, and Business Cycles, The MIT Press, Cambridge, Massachussetts.

Cuthbertson, K., S. Hall and M. Taylor (1992), Applied Econometric Techniques, Philip Allan, London.

Daniel, W. (1990), Applied Nonparametric Statistics, 2nd edition, PWS-Kent, Boston.

Darby, J., S. Wren-Lewis (1992), "Changing trends in international manufacturing productivity", Scandinavian Journal of Economics, 34, 3, 457-477.

De Grauwe, P. (1994), Economics of Monetary Integration, Oxford University Press, Oxford.

DeJong, N. D., J. Nankernis, N. E. Sarin and C. Whiteman (1992), "Integration vs. trend stationarity in time series", Econometrica, 60, 2, March, 423-433.

De Long, J. (1988), "Productivity growth, convergence, and welfare: comment", American Economic Review, 78, 5, December, 1138-1154.

Deutsch, J. (1992), "Linear regression under two separate regimes: an empirical distribution for Quandt's log likelihood ratio", Applied Economics, 24, 1, January, 123-127.

Dollar, D., E. Wolf and W. Baumol (1988), "The factor-price equalisation model and industry labor productivity: an empirical test across countries", in R. Feensta (1988), 23-47.

Dowrick, S. (1992), "Technological catch up and diverging incomes: patterns of economic growth 1960-88", Economic Journal, 102, 412, May, 600-610.

Dowrick, S. and N. Gemmell (1991), "Industrialization, catching up and economic growth: a comparative study across the world's capitalist economies", Economic Journal, 101, 405, March, 263-275.

Dowrick, S. and D.-T. Nguyen (1989), "OECD comparative economic growth 1950-85: catch up and convergence", American Economic Review, 79, 5, December, 1010-30.

Durlauf, S. and P. Johnson (1992), Local Vs. Global Convergence Across National Economies, NBER working paper no. 3996, Cambridge, Massachusetts.

Easterly, W., M. Kremer, L. Pritchett and L. Summers (1993), "Good policy or luck? Country growth performance and temporary shocks", Journal of Monetary Economics, 32, 3, 459-483.

Eichengreen, B. (1993), "European monetary unification", The Journal of Economic

Literature, 31, 3, September, 1321-1357.

Engle, R. and W. Granger, ed., (1991), Long-run Economic Relationships, Oxford University Press, Oxford.

Feensta, R. C., ed., (1988), Empirical Methods for International Trade, The MIT Press, Cambridge, Massuchussetts.

Formby, T., R. C. Hill and S. R. Johnson (1984), Advanced Econometric Methods, Springer-Verlag, New York.

Fraser, P., C. V. Helliar and D. M. Power (1994), "An empirical investigation of convergence among European equity markets", Applied Financial Economics, 4, 2, April, 149-157.

Friedman, M. (1992),"Do old fallacies ever die?", Journal of Economic Literature, 30, 4, December, 1010-1030.

Gremmen, H. (1985), "Testing the factor price equalisation theorem in the EC: An alternative approach", Journal of Common Market Studies, 23, 1, March, 277-286.

Grier, K. and G. Tullock (1989), "An empirical analysis of cross-national economic growth", Journal of Monetary Economics, 24, 2, September, 259-276.

Gros, D. and N. Thygesen (1992), European Monetary Integration, From the European Monetary System to European Monetary Integration, Longman, Harlow, Essex.

Grossman, G. and E. Helpman (1991a), "Quality ladders in the theory of growth", Review of Economic Studies, 58, 193, 43-61.

Grossman, G.and E. Helpman (1991b), "Trade, knowledge spillovers, and growth", European Economic Review, 35, 2/3, April, 517-526.

Grossman, G. and E. Helpman (1994), "Endogenous innovation in the theory of growth", Journal of Economic Perspectives, 8, 1, 23-44.

Haldane, A. and S. Hall (1991), “Sterling’s relationship with the dollar and the deutschmark: 1976-89”, Economic Journal, 101, 406, May, 436-443.

Hall, S., D. Robertson and M. Wickens (1992), Measuring Convergence of the EC Economies, CEF discussion paper no. 01-92, London Business School, London.

Hall, S., D. Robertson and M. Wickens (1993), How to Measure Convergence with a Application to the EC Economies, CEF discussion paper no. 19-93, London Business School, London.

Hammond, P. J. and A. Rodriguez-Clare (1993), “On endogeneizing long-run growth”, Scandinavian Journal of Economics, 95, 4, 391-425.

Harvey, A. (1989), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge.

Helpman, E. (1992), “Endogenous macroeconomic growth theory”, in Cukierman, A., Z. Hercowitz, and L. Leiderman (1992), 83-115.

Honohan, P. (1992), “Price and monetary convergence in currency unions: the Franc and Rand Zones”, Journal of International Money and Finance, 11, 4, August, 397-410.

Hulten, C. and R. Schwab (1993), Endogenous Growth, Public Capital, and the Convergence of Regional Manufacturing Industries, NBER working paper no. 4538, Cambridge, Massachusetts.

Inada, K.-I. (1963), “On a two sector model of economic growth: comments and a generalization”, Review of Economic Studies, 30, June, 119-27.

Janacek, G. and L. Swift (1993), Time Series, Forecasting, Simulation, Applications, Ellis Horwood, London.

Jonavic, B. and S. Lach (1991), The Diffusion of Technology and Inequality Among Nations, NBER working paper no. 3732, Cambridge, Massachusetts.

Jones, L. and R. Manuelli (1990), "A convex model of equilibrium growth: theory and policy implications", Journal of Political Economy, 98, 5, 485-517.

Karfakis, C. J. and D. M. Moschos (1990), "Interest rates and the European Monetary System: a time series analysis", Journal of Money, Credit and Banking, 22, 3, 388-394.

Kelly, M. (1992), "On endogenous growth with productivity shocks", Journal of Monetary Economics, 30, 1, 47-56.

Kemeny, J. G. and J. L. Snell (1960), Finite Markov Chains, D. Van Nostrand Company, Princeton, New Jersey.

Kim, S. and H. Mohtadi (1992) "Labor specialization and endogenous growth", American Economic Review, 82, 2, May, 404-408.

King, R., C. Plosser, J. Stock and M. Watson (1991), "Stochastic trends and economic fluctuations", American Economic Review, 81, 4, September, 819-840.

King, R. and S. Rebelo (1990), "Public policy and economic growth: Developing neoclassical implications", Journal of Political Economy, 98, 5 (part 2), S126-S150.

King, R. and S. Rebelo (1993), "Transitional dynamics and economic growth in the neoclassical model", American Economic Review, 83, 4, September, 908-931.

Knight, M., N. Loyasa and D. Villanueva (1993), "Testing the neoclassical theory of economic growth", IMF Staff Papers, 40, 3, September, 512-541.

Koopmans, T. C. (1965), "On the concept of optimal economic growth", in Pontificiae Academiae Scientiarum Scripta Varia: Study Week on the Econometric Approach to

Development Planning, 28, Rand McNally, Chicago.

Lee, T., G. Judge and A. Zellner (1970), Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, North-Holland, Amsterdam.

Levine, R. and D. Renelt (1992), "A sensitivity analysis of cross-country growth regressions", American Economic Review, 82, 4, September, 942-963.

Lichtenberg, F. R. (1994), "Testing the convergence hypothesis", The Review of Economic and Statistics, 76, 3, August, 576-579.

Lucas, R. (1988), "On the mechanics of economic development", Journal of Monetary Economics, 22, 1, July, 3-42.

Lucas, R. (1990), "Why doesn't capital flow from rich to poor countries?", American Economic Review, 80, 2, May, 92-96.

Lucas, R. (1993), "Making a miracle", Econometrica, 61, 2, March, 251-272.

MacKinnon, J. G. (1991), "Critical values for cointegration tests", in Engle and Granger (1991), eds., 267-276.

Maddala, G. S. (1992), Introduction to Econometrics, 2nd edition, MacMillan Publishing Company, New York.

Maddison, A. (1979), "Per capita output in the long-run", Kiklos, 39, 1/2, 412-429.

Maddison, A. (1989), The World Economy in the 20th Century, Development Centre Studies, OECD, Paris.

Maddison, A. (1991), Dynamic Fases in Capitalist Development, Oxford University Press, Oxford.

Mankiw, N. G., D. Romer and D. N. Weil (1992), "A contribution to the empirics of economic growth", Quarterly Journal of Economics, 107, 2, May, 407-37.

Mokhtari, M. and F. Rassekh (1989), "Tendency towards factor price equalisation among OECD countries", Review of Economics and Statistics, 4, 71, November, 636-642.

Nelson, R. (1991), "Diffusion of development: post-World War II convergence among industrial nations", American Economic Review, 81, 2, May, 271-275.

Neven, D. and C. Gouyette (1995), "Regional convergence in the European Community", Journal of Common Market Studies, 33, 1, March, 47-65.

Ochrane, J. (1991), "A critique of the application of unit root tests", Journal of Economic Dynamics and Control, 15, 2, April, 275-284.

Ogaki, M. (1992), "Growth in open economies: a comment", Carnegie-Rochester Conference Series on Public Policy, 36, 47-55.

Pack, H. (1994), "Endogenous growth theory: intellectual appeal and empirical shortcomings", Journal of Economic Perspectives, 8, 1, Winter, 55-72.

Parente, S. and E. Prescott (1991), Technology Adoption and Growth, NBER working paper no. 3733, Cambridge, Massachusetts.

Pesaran, B. and M. H. Pesaran (1991), Microfit 3.0, An Interactive Econometric Software Package, User Manual, Oxford University Press, Oxford.

Pigott, C. A. (1994), "International interest rate convergence: a survey of the issues and evidence", Federal Reserve Bank of New York Quarterly Review, 18, 4, Winter, 24-37.

Quah, D. (1992), International Patterns of Growth: II. Persistance, Path Dependence, and Sustained Take-off in Growth Transition, mimeo, LSE Economics Department, October.

Quah, D. (1993a), "Empirical Cross-Section Dynamics in Economic Growth", European Economic Review, 37, 2/3, April, 426-434.

Quah, D. (1993b), "Galton's fallacy and tests of the convergence hypothesis", Scandinavian Journal of Economics, 95, 4, 427-443.

Quah, D. (1993c), Dependence in Growth and Fluctuations Across Economies with Mobile Capital, mimeo, LSE Economics Department, June.

Quah, D. (1993d), Exploiting Cross section Variation for Unit Root Inference in Dynamic Data, mimeo, LSE Economics Department, June.

Quah, D. (1994), Empirics for Economic Growth and Convergence, mimeo, LSE Economics Department, November.

Quandt, R. (1958), "The estimation of the parameters of a linear regression system obeying two separate regimes", Journal of the American Statistical Association, 53, 873-880.

Quandt, R. (1960), "Test of the hypothesis that a linear regression obeys two separate regimes", Journal of the American Statistical Association, 55, 324-330.

Rebelo, S. (1991), "Long-run policy analysis and long-run growth", Journal of Political Economy, 99, 3, June, 500-521.

Rebelo, S. (1992), "Growth in open economies", Carnegie-Rochester Conference Series on Public Policy, 36, 5-46.

Rivera-Batiz, L. and D. Xie (1992), "GATT, trade, and growth", American Economic Review, 82, 2, May, 422-427.

Rivera-Batiz, L. and P. Romer (1991a), "Economic integration and endogenous growth", Quarterly Journal of Economics, 106, 2, May, 531-555.

Rivera-Batiz, L. and P. Romer (1991b), "International trade with endogenous technological change", European Economic Review, 35, 4, May, 971-1004.

Romer, P. (1986), "Increasing returns and long-run economic growth", Journal of Political Economy, 95, 5, October, 1002-37.

Romer, P. (1990a), "Endogenous technological change", Journal of Political Economy, 98, 5, October, S71-S102.

Romer, P. (1990b), "Are nonconvexities important for understanding growth?", American Economic Review, 80, 2, March-May, 86-91.

Romer, P. (1991), "Increasing returns and new developments in the theory of growth", in Barnett, Cornett, D'Aspremont, Gabsewickz and Mas-Colell, eds. (1991), 83-110.

Romer, P. (1994), "The origins of endogenous growth", Journal of Economic Perspectives, 8, 1, Winter, 3-22.

Sala-i-Martin, X. (1994), "Cross-sectional regressions and the empirics of economic growth", European Economic Review, 38, 3/4, April, 739-747.

Samuelson, P. (1948), "International trade and the equalisation of factor prices", The Economic Journal, 58, 163-184 (reproduced in Stiglitz, ed., (1966), 2nd volume, 847-868).

Samuelson, P. (1949), "International factor price equalisation once again", The Economic Journal, 59, 869-885 (reproduced in Stiglitz, ed., (1966), 2nd volume, 869-884).

Scott, M. (1993), "Exploring economic growth", American Economic Review, 83, 2, May, 421-425.

Shaw, G. K. (1992), "Policy implications of endogenous growth theory", Economic Journal, 102, 412, May, 611-62.

Shell, K., ed., (1967), Essays on the Theory of Optimal Economic Growth, The MIT Press, Cambridge, Massachussetts.

Sheshinski, E. (1967), "Optimal accumulation with learning by doing", in Shell, K., ed., (1967), 31-52.

Silverman, B. W. (1986), Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York.

Solow, R. (1957), "Technical change and the aggregate production function", Review of Economics and Statistics, 39, August, 312-320

Stiglitz, J (1966), ed., The Collected Scientific Papers of Paul A. Samuelson, 3 volumes, The MIT Press, Cambridge, Massachussetts.

Stockman, A. C. (1988), "Sectoral and national aggregate disturbances to industrial output in seven European countries", Journal of Monetary Economics, 21, March-May, 287-409.

Summers, R. and A. Heston (1988), "A new set of international comparisons of real product and price levels estimates for 130 countries, 1950-1988", Review of Income and Wealth, 34, 1, 1-25.

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, 2, 327-368.

Tamura, R. (1991), "Income convergence in an endogenous growth model", Journal of Political Economy, 99, 3, June, 522-540.

Tovias, A. (1982), "Factor price equalisation in the EEC", Journal of Common Market Studies, 20, 4, June, 375-388.

Verspagen, B. (1992), "Endogenous innovation in neo-classical growth models: a survey", Journal of Macroeconomics, 14, 4, Fall, 631-662.

Von Hagen, J. and M. Neumann (1992), Real Exchange Rates within and Between Currency Areas: How Far is Emu?, CEPR discussion paper no. 660, London.

Young, A. (1993), "Invention and bounded learning by doing", Journal of Political Economy, 101, 3, June, 443-472.

Zagardo, J. T. (1991), "Worldwide convergence of productivity levels: recent empirical evidence", Business Economics, 26, 4, October, 19-25.