Introduction to the Empirical Analysis of Economic Growth

Mario Lavezzi

lavezzi@unipa.it

International Doctoral Program in Economics, SSSUP
and
Ph.D in Economics,University of Pisa

October 2009

The issues

- Theoretical issues
  1. The shape of the growth process (convergence)
     - Solow model
     - Endogenous growth (AK model)
     - Multiple equilibria model (nonlinearities)
  2. Identification of growth determinants
  3. Parameter heterogeneity. Typical question: adding one "unit" of human capital in the U.S. or in Ghana has the same marginal effect on growth? (nonlinearities)
- Empirical issues
  1. Which method for the empirical investigation? Growth regressions, distribution dynamics, cluster analysis, etc.
  2. Which variables to consider? (model uncertainty)

Reading list

Azariadis, C. and A. Drazen (1990), "Threshold Externalities in Economic Development", *Quarterly Journal of Economics* 105, 501-26.

Barro, R. J. (1991), "Economic Growth in a Cross Section of Countries", *Quarterly Journal of Economics* 106, 407-43

Barro, R. J. and X. Sala-i-Martin (2004), "Economic Growth. Second Edition", MIT Press.

Bernard, A. and S. N. Durlauf (1996), "Interpreting Tests of Convergence Hypothesis", *Journal of Econometrics* 71, 161-73.

Bowman, A. W. and A. Azzalini (1997), "Applied Smoothing Techniques for Data Analysis", Clarendon Press.

Durlauf, S. N. and P. A. Johnson (1995), "Multiple Regime and Cross-Country Growth Behaviour", *Journal of Applied Econometrics* 10, 365-84.

Durlauf, S. N., P. A. Johnson and J. R. W. Temple (2005), "Growth Econometrics", in S. N. Durlauf and P. Aghion (Eds.), Handbook of Economic Growth, Elsevier.

Durlauf, S. N., A. Kourtellos and A. Minkin (2001), "The Local Solow Growth Model", *European Economic Review* 45, 928-40.

Durlauf, S. N. and D. Quah (1999), "The New Empirics of Economic Growth" in J. Taylor and M: Woodford (Eds.), *Handbook of Macroeconomics*, Noth Holland.

Fiaschi, D. and A. M. Lavezzi (2003), "Distribution Dynamics and Nonlinear Growth", *Journal of Economic Growth* 8, 379-401.

Galor, O. (1996), "Convergence? Inferences from Theoretical Models", *Economic Journal* 106, 1056-69 (and the Symposium in the same issue)

Härdle, W., M. Müller, S. Sperlich and A. Werwatz (2004) "Nonparametric and Semiparametric Models", Springer.

Isaacson, D. L. and R. W. Madsen (1976), *Markov Chains: Theory and Applications*. New York: Wiley.

Johnson, P. A. (2005), "A Continuous State Space Approach to 'Convergence by Parts'", *Economics Letters* 86, 317-21.

Jones, C. I. (1997), "On the Evolution of World Income Distribution", *Journal of Economic Perspectives* 11, 19-36

Kalaitzidakis, P., T. Mamuneas, A. Savvides and T. Stengos (2001), "Measures of Human Capital and Nonlinearities in Economic Growth", *Journal of Economic Growth* 6, 229-54.

Liu, Z. and T. Stengos (1999). "Non-linearities in Cross-Country Growth Regressions: a Semiparametric

Approach", *Journal of Applied Econometrics* 14, 527-38.

Maddison A. (2001), *The World Economy: a Millenium Prospective*. Paris: OECD.

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

Quah, D. T. (1993), "Empirical Cross-section Dynamics for Economic Growth", *European Economic Review* 37, 426-34.

Rostow, W. W. (1960), *The Stages of Economic Growth*. Cambridge University Press

Solow, R. M. (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics* 70, 65-94.

Temple, J. (1999), "The New Growth Evidence", *Journal of Economic Literature* 37, 112-56.

Convergence

(see Barro and Sala-i-Martin, 2004, Ch. 1, and Durlauf et al., 2005, Section 4)

- Starting question: will poor countries catch up with the richest?

- Studying convergence is an instance of the study of: "long-run outcomes from contemporary behaviors" (Durlauf et al., 2005, p. 559)

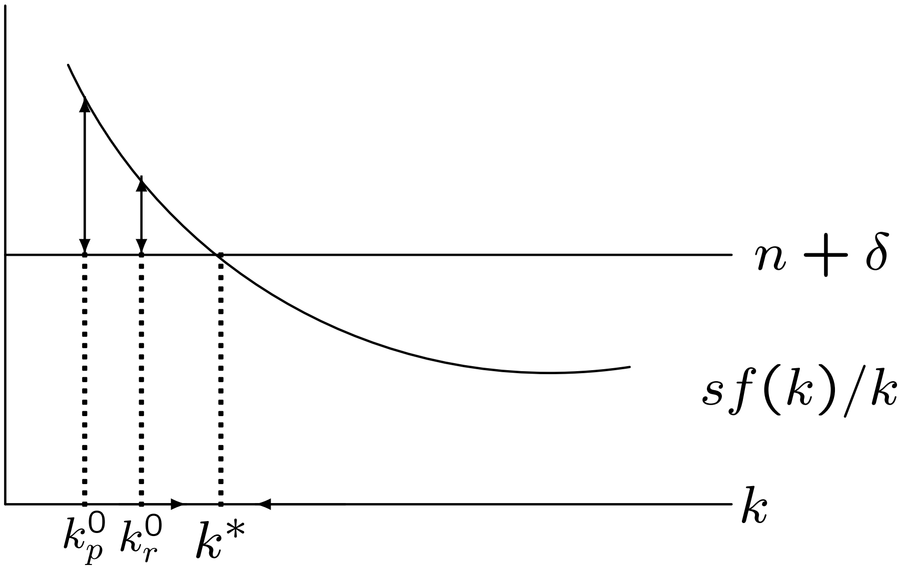
- Different predictions on convergence are implied by different growth models.
  - Absolute convergence (Solow, 1956)
  - Conditional convergence (Solow, 1956)
  - No convergence (divergence) (Endogenuous growth)
  - Club convergence (e. g. Azariadis and Drazen, 1990)

- Different concepts of convergence
  - β-convergence
  - σ-convergence

The Solow model: absolute convergence

$$\dot{k} = sy - (n+\delta)k = sf(k) - (n+\delta)k$$

$$\frac{\dot{k}}{k} = \gamma_k = \frac{sf(k)}{k} - (n+\delta)$$



- Absolute convergence: per capita income of countries converge to one another in the long run, independently of their initial conditions (Galor, 1996)

- Poor countries grow faster than rich countries (in the transition)

- Transitory shocks on capital/income have not permanent effects

- The hypothesis of absolute convergence is typically rejected in cross-country analyses.

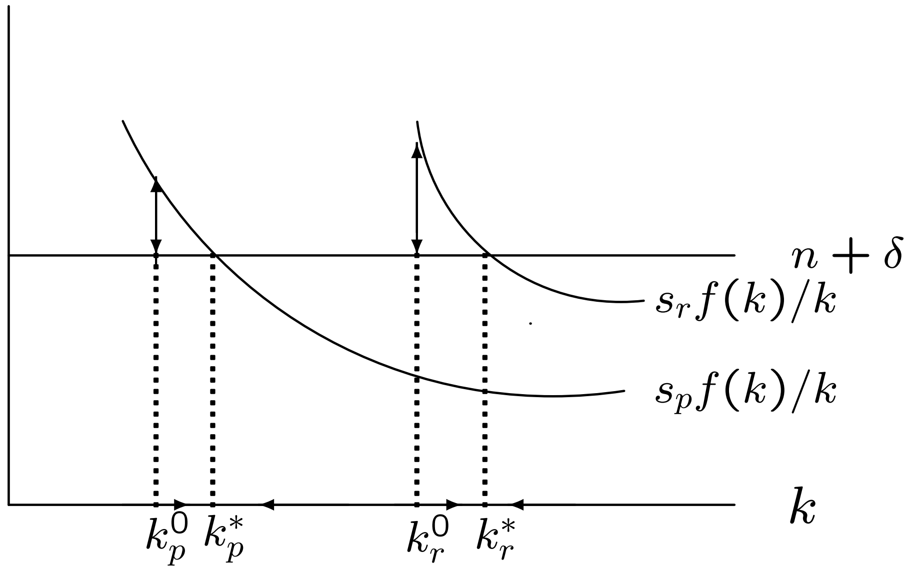
Figure 1 from Temple (1999):

– “Triangular” shape of the relation growth rate - initial income

– Not all poor countries grow faster than rich countries

– Differences in income are expected to persist

The Solow Model: conditional convergence



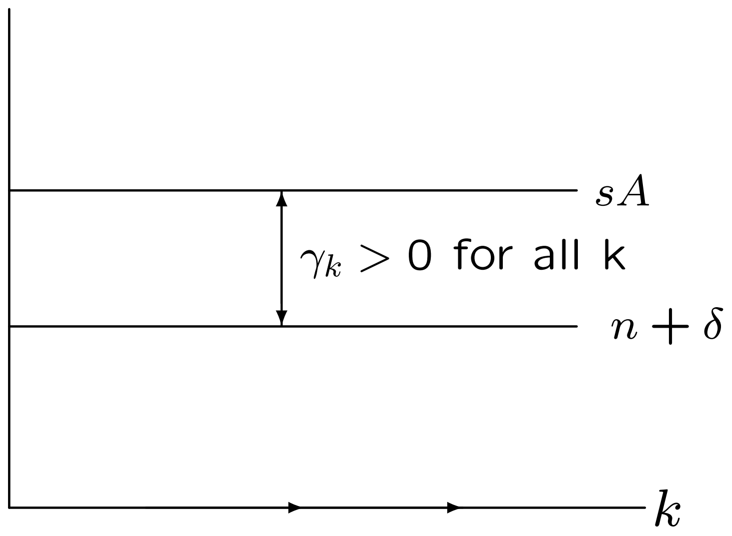
- Conditional convergence: per capita incomes of countries that are identical in their structural characteristics (e.g. preferences, technologies, rates of population growth, government policies, etc.) converge to one another in the long run independently of their initial conditions (Galor, 1996)
- Rich countries can grow faster than poor countries (in the transition)
- Transitory shocks on capital/income have not permanent effects
- The rejection of the hypothesis of absolute convergence does not imply the rejection of the Solow Model
- Observing persistent differences in income requires an explanation of persistent differences in structural parameters

AK Model (endogenous growth)

$$Y = AK$$

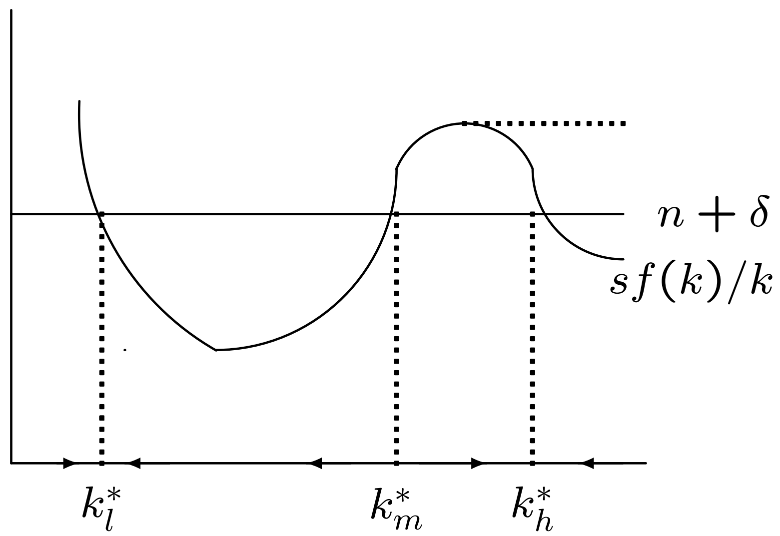
$$y = Ak$$

$$\frac{\dot{k}}{k} = \gamma_k = sA - (n + \delta)$$



- No transitional dynamics
- No absolute convergence
- No conditional convergence
- Transitory shocks on capital/income have permanent effects
- No strong support for the AK model

Multiple equilibria model



- Club convergence (polarization, persistent poverty and clustering): per capita incomes of countries that are identical in their structural characteristics converge to one another in the long run provided that their initial conditions are similar as well, e. g. they are in the same basin of attraction (Galor, 1996)

- Rich countries can grow faster than poor countries (in the transition)

- Transitory shocks on capital/income have permanent effects

- No need to explain why economies remain structurally different

- Structurally similar economies can show persistent differences in income

- Structurally similar economies with similar initial conditions may diverge!

- The equilibrium level k_l^* is a poverty trap: stable steady state level of cpaital (and income)

- The issue is the existence of an intermediate range of capital in which the relation growth rate/capital level is increasing

- Possible explanations:

  – Technological spillovers: after a threshold level of capital, the average product of capital grows with k. In other words, technological progress depends on the stock of physical (and human) capital (Ex. in Azariadis and Drazen, 1990, technological externalities with a threshold property: discontinuity in the aggregate production function)

  – Structural transformation of the economy (Rostow, 1960): in early stages (low k), the economy is essentially based on agriculture (subject to diminishing returns), then it industrializes (take-off), then it reaches a stage of maturiry.

Two concepts of convergence

(Barro and Sala-i-Martin, 2004, pp. 50-51)

1. β-convergence:

   (a) unconditional: economies with lower levels of per capita income tend to grow faster in per capita terms

   (b) conditional: economies with lower levels of per capita income (expressed relative to their steady-state levels of per capita income) tend to grow faster in per capita terms

2. σ-convergence: the dispersion of real per capita income across a group of economies tends to fall over time. That is, σ-convergence holds between times t and $t + T$ if:

   $$D_{\log y,t} > D_{\log y,t+T}$$

   the sample dispersion of (log) incomes decreases over time.

• <u>β-convergence does not imply σ-convergence!</u>

Consider:

$$log(y_{i,t}) = a + (1-b)log(y_{i,t-1}) + u_{i,t}$$

with $0 < b < 1$ implies absolute convergence, as annual growth rate, $log(y_{i,t}/y_{i,t-1})$ is inversely related to $log(y_{i,t-1})$. $u_{i,t}$: disturbance term, zero mean, variance given by σ_u^2.

Sample variance, used to measure dispersion, follows:

$$D_t \approx (1-b)^2 D_{t-1} + \sigma_u^2$$

which implies a steady state dispersion equal to:

$$D^* = \sigma_u^2/[1-(1-b)^2]$$

Even if $b > 0$, $D^* > 0$ as long as $\sigma_u^2 > 0$ (i.e. beta-convergence does not, in any case, imply reduction of sample dispersion to zero)

The evolution of D_t can be expressed as: (substitute D_t with $D_t - D^*$)

$$D_t = D^* + (1-b)^{2t} \cdot (D_0 - D^*)$$

we have that, since $0 < b < 1$, D_t monotonically approaches the steady-state value D^*, in particular:

- if $D_0 > D^*$, D_t falls over time
- if $D_0 < D^*$, D_t increases over time

Hence, $\underline{D_t \text{ can rise over time even if } b > 0}$

Methods of empirical analysis

- Linear regression
- Distribution dynamics
- Other

Linear growth regressions

Three types of equations to estimate:

1. $\gamma_i = \alpha + \beta y_{i,0} + \epsilon_i$

   (where $\gamma_{i,T}$ is the average annual growth rate of per capita income between, say, 0 and T.)

   If $\widehat{\beta} < 0 \rightarrow$ absolute convergence - unconditional β-convergence

2. $\gamma_i = \alpha + \beta y_{i,0} + \psi \mathbf{X}_i + \epsilon_i$

   if $\widehat{\beta} < 0 \rightarrow$ conditional convergence - conditional β-convergence

   $y_{i,0}$ and $\mathbf{X}_i$: Solow growth determinants (saving rate, population growth rate, rate of obsolescence of capital).

3. $\gamma_i = \alpha + \beta y_{i,0} + \psi \mathbf{X}_i + \pi \mathbf{Z}_i + \epsilon_i$

   if $\widehat{\beta} < 0 \rightarrow$ conditional convergence - conditional β-convergence

   "Barro regression". $\mathbf{Z}_i$: growth determinants not in Solow's model (democracy, education, financial development, ...).

Some remarks

1. General comment: $\mathbf{X}_i$ and $\mathbf{Z}_i$ are typically referred to as indicators of structural heterogeneity, which would imply conditional convergence, as something different from the effect of initial conditions, which could include the stock of physical and human capital, etc. One problem: the variables taken as proxy for structural characteristics may be endogenously determined by initial conditions (ex. low income $\rightarrow$ low level of democracy).

2. A remark on β-convergence

   - β-convergence:

     $\widehat{\beta} < 0$ without controls $\rightarrow$
     unconditional β-convergence

     $\widehat{\beta} < 0$ with controls $\rightarrow$ conditional β-convergence

   - Bernard and Durlauf (1996) show that:

   (a) the coefficient $\widehat{\beta}$ is a weighted average. Some countries in the sample may follow the Solow model, some may not.

   (b) $\widehat{\beta} < 0$ is not sufficient to conclude that there is β-convergence. The data can be generated from a model with multiple equilibria but the regression on the misspecified model can nonetheless return a negative $\widehat{\beta}$. In the sample, some countries may be converging some may not: "the test is ill-designed to analyze [this]" (Bernard and Durlauf, 1996, p. 167)

The Distribution Dynamics Approach

(see Durlauf et al., 2005, Sect. 4.3)

• Starting point: to analyze the evolution of the whole income distribution

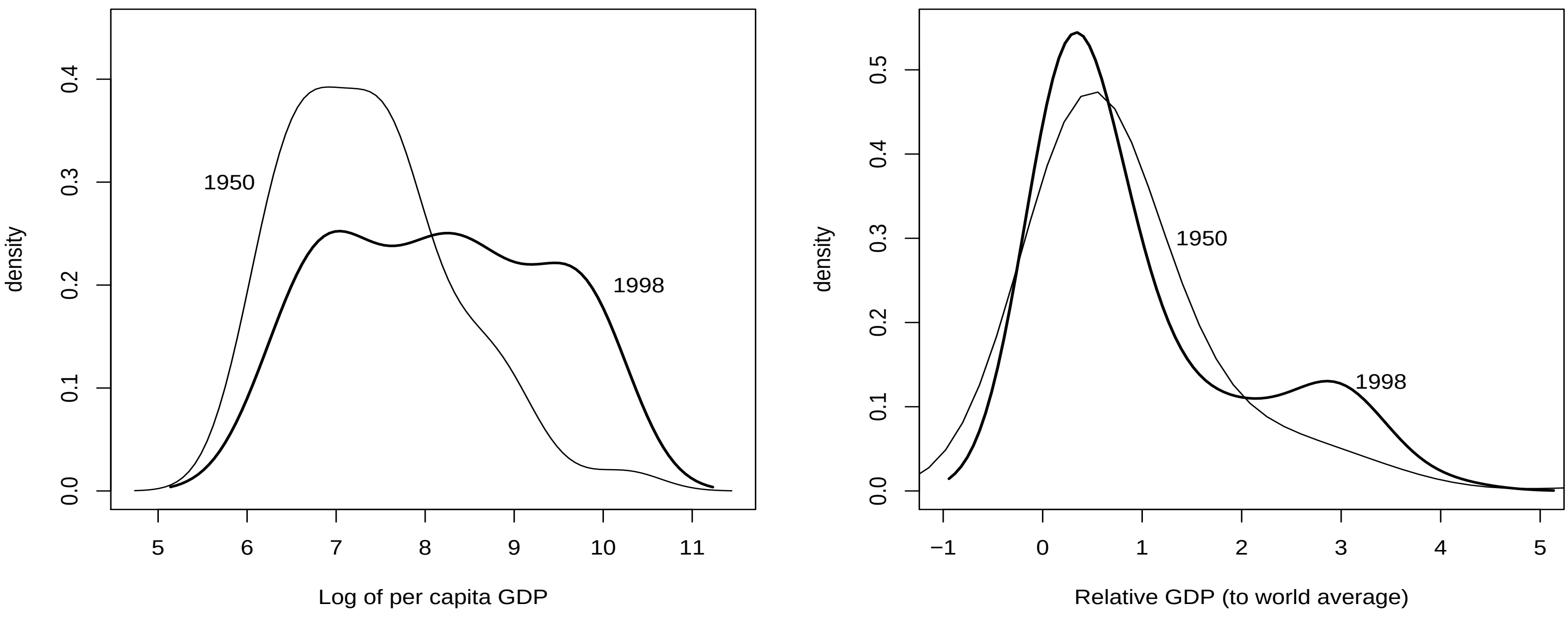
• Motivation: dissatisfaction with the standard approach based on cross-section regressions

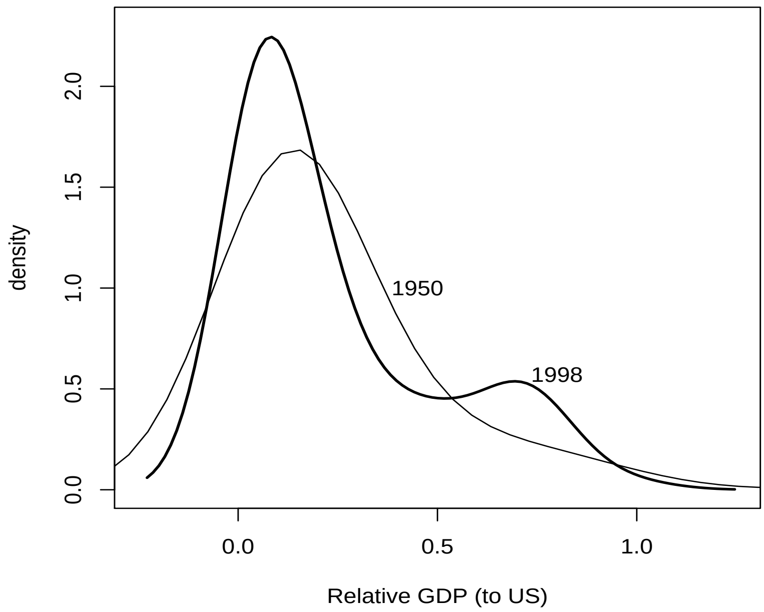
• This approach is more informative on convergence, divergence, intradistribution dynamics, catching up and falling behind (see Figure 1, p. 99, from Durlauf and Quah, 1999).

- <u>An example</u>

- Data from Maddison (2001), 122 countries, 1950-1998

- Estimates of distribution of per capita GDP in 1950 and in 1998





- Note the emergence of “twin peaks”

- This is considered as evidence of polarization: the distribution becomes thinner in the center and thicker in the tails

- Remark on data: i) absolute values; ii) relative to sample average; iii) relative to US

• A possible way to look at the evolution of the cross-country income distribution is to represent it as a Markov Chain

• The cross-country distribution is represented by a vector: any element is an income interval and its value indicates the proportion of countries in that interval in a given period

• Example: 3 income levels. State space of the process: $S = (1, 2, 3)$

$\mathbf{q}_t = [q_{1,t}, q_{2,t}, q_{3,t}]$, $0 \leq q_{1,t}, q_{2,t}, q_{3,t} \leq 1$,

$\sum_{i=1}^{3} q_{i,t} = 1$

Dynamics is given by: $\mathbf{q}_{t+1} = \mathbf{q}_t \mathbf{P}$

where **P** is a transition matrix:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

$0 \leq p_{ij} \leq 1$, $\sum_{j=1}^{3} p_{ij} = 1$, $\forall i$

- An element of $\mathbf{P}$ is a transition probability. It is a conditional probability, for example:

  $p_{11} = P(X_{t+1} = 1 | X_t = 1)$

  where X_t is the state of the process at time t, i.e. the income class of a country at time t

- Markov property: the state of the process at time $t+1$ only depends on the state of the process at time t, and not on other past periods, e. g. we do not have that:

  $p_{..1,1} = P(X_{t+1} = 1 | X_t = 1, X_{t-1} = ...,$
  $X_{t-2} = ..., ...)$

- In the present case the Markov Chain is stationary, that is the transition matrix is the same in every period. If the process is non-stationary, the transition matrix would be indexed by t, $\mathbf{P_t}$

- Long run dynamics:

  $\mathbf{q}_1 = \mathbf{q}_0\mathbf{P}$

  $\mathbf{q}_2 = \mathbf{q}_1\mathbf{P} = \mathbf{q}_0\mathbf{P}^2$

  ...

  $\mathbf{q}_n = \mathbf{q}_0\mathbf{P}^n$

  Under some regularity conditions:

  $\overline{\mathbf{q}} = \overline{\mathbf{q}}\mathbf{P}$

  and the process is *ergodic*.

- $\overline{\mathbf{q}}$ is defined as *stationary*, *invariant* or *ergodic* distribution of the process.

A numerical example. The transition matrix:

$$\mathbf{P}=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.6 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{bmatrix}$$

is associated to the invariant distribution:

$\overline{\mathbf{q}} = [0.4146, 0.4024, 0.1829]$

• <u>From Quah (1993)</u>

Data on real GDP per capita (relative to world average)

$\#obs$	1/4	1/2	1	2	∞
456	0.97	0.03	0	0	0
643	0.05	0.92	0.04	0	0
639	0	0.04	0.92	0.04	0
468	0	0	0.04	0.94	0.02
508	0	0	0	0.01	0.99
$Ergodic$	0.24	0.18	0.16	0.16	0.27

• Results by Quah: emergence of twin peaks

• A more optimistic view is in Jones (1997)

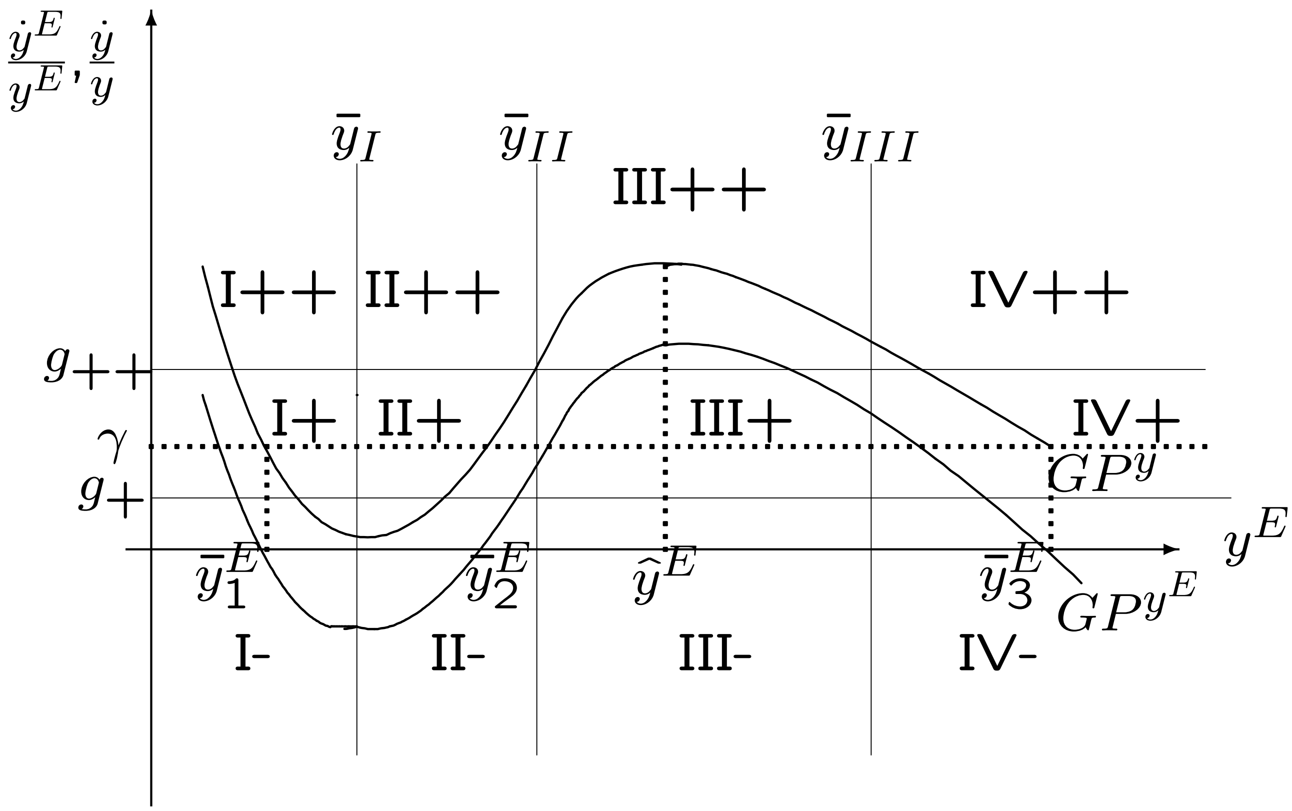
• Empirically, transition probabilities can be estimated by frequencies of transitions:

$$\widehat{p_{ij}} = \frac{n_{ij}}{n_i}$$

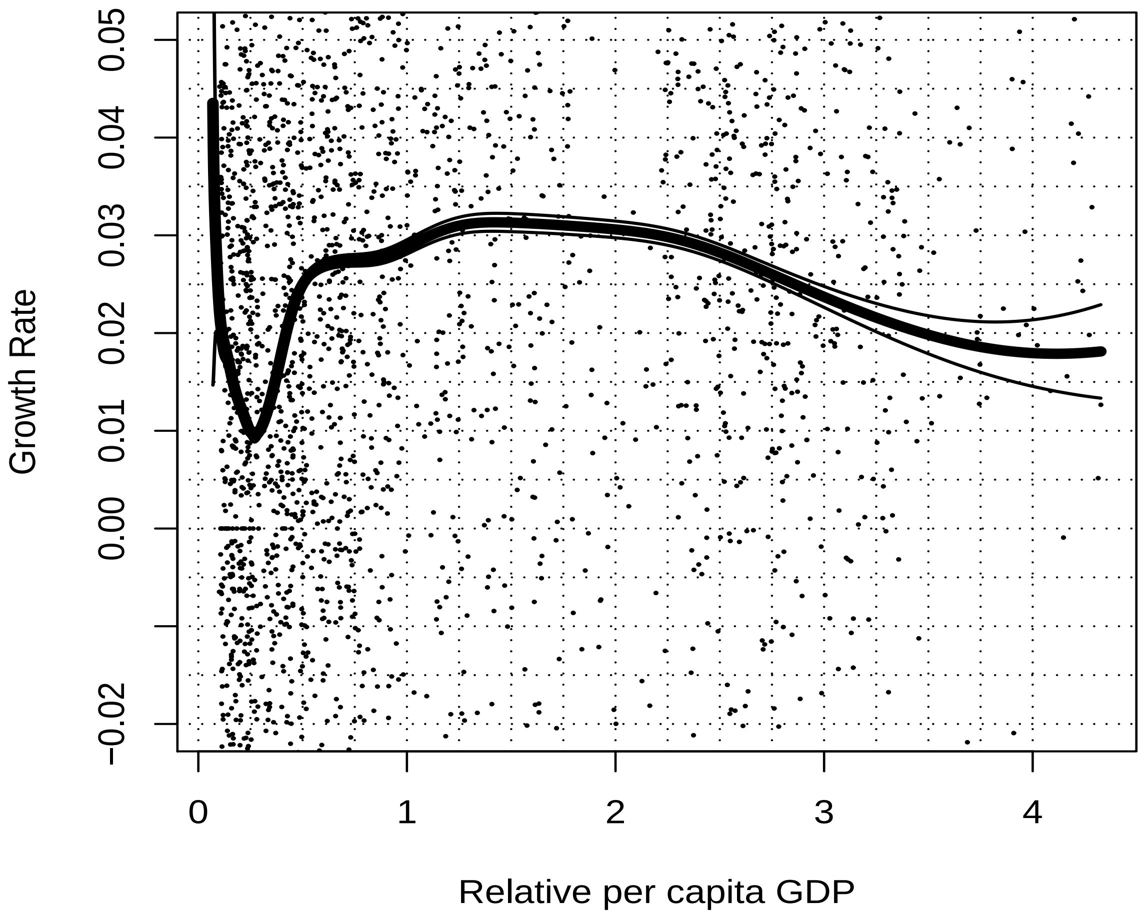
where n_{ij} is the number of transitions from state i to state j and n_i is the number of observations in state i. These estimates are the maximum likelihood estimates of the true (unknown) transition probabilities.

• A possible extension of the distribution dynamics approach is in Fiaschi and Lavezzi (2003)

• Idea: to extend the approach to the study of the shape of the growth process: the state space is defined in terms of income levels and growth rates.



• First step: run a nonparametric regression of growth rates against income levels



• Second step: study the distribution dynamics

- Definition of the state space

Income\Growth rate	$< 0.8\%$	$0.8\% - 2.8\%$	$> 2.8\%$
$0 - 0.3\mu_I$	I-	I+	I++
$0.3\mu_I - 0.9\mu_I$	II-	II+	II++
$0.9\mu_I - 2.5\mu_I$	III-	II+	III++
$> 2.5\mu_I$	IV-	IV+	IV++

- Transition matrix

Obs	States	I-	I+	I++	II-	II+	II++	III-	III+	III++	IV-	IV+	IV++
423	I-	0.54	0.14	0.32	0	0	0	0	0	0	0	0	0
118	I+	0.42	0.20	**0.37**	0	0	0	0	0	0	0	0	0
337	I++	0.39	0.12	**0.45**	0.02	0	0.01	0	0	0	0	0	0
470	II-	0.03	0.01	0.01	0.47	0.14	0.34	0	0	0	0	0	0
221	II+	0	0	0	0.35	0.22	**0.42**	0	0	0	0	0	0
593	II++	0	0	0	0.26	0.16	**0.53**	0.01	0	0.04	0	0	0
202	III-	0	0	0	0.06	0.01	0.04	0.46	0.16	0.26	0	0	0
132	III+	0	0	0	0	0.01	0	0.23	0.17	**0.55**	0.01	0.02	0.02
445	III++	0	0	0	0	0	0	0.16	0.16	**0.65**	0	0	0.02
93	IV-	0	0	0	0	0	0	0.05	0.02	0.02	0.29	0.30	0.31
125	IV+	0	0	0	0	0	0	0	0.02	0.01	0.23	0.34	**0.39**
201	IV++	0	0	0	0	0	0	0	0	0	0.16	0.27	**0.56**

• Distribution dynamics and ergodic distribution

	I	II	III	IV
1960	0.20	0.47	0.22	0.11
1989	0.31	0.34	0.19	0.16
Ergodic	0.41	0.28	0.18	0.14

• Normalized ergodic distribution

	-	+	++
I	0.48	0.14	0.38
II	0.38	0.17	0.45
III	0.27	0.17	0.56
IV	0.22	0.31	0.47

Distribution dynamics with continuous state space

(See also Johnson, 2005, and Johnson's web-page)

• Problem: discretization of state space may distort the underlying dynamics, especially in the long run

• Possible solution: avoid discretization

• Repeat the analysis with continuous GDP space:

• $f_t(y)$: density of cross-country income distribution at time t

• $f_{t+\tau}(y)$: density of cross-country income distribution at time $t + \tau$

• Under the assumptions: i) the transition process is time-invariant; ii) the process is first-order, we can write the distribution dynamics as:

$$f_{t+\tau}(x) = \int_0^\infty g_\tau(x|z) f_t(z) dz$$

where $x = y_{t+\tau}$, $z = y_t$.

• $g_\tau(x|z)$ is the τ-period ahead density of x conditional on z

• it is the continuous analogue of a transition matrix. It maps the distribution of time t into the distribution at time $t + \tau$

• it is defined stochastic kernel (see Figures 11a and 11b, from Durlauf and Quah, 1999)

• If the ergodic distribution implied by $g_\tau(x|z)$ exists, $f_\infty(x)$, it satisfies:

$$f_\infty(x) = \int_0^\infty g_\tau(x|z) f_\infty(z) dz$$

On nonparametric models

(see Härdle et. al, 2004, and Bowman and Azzalini, 1997)

1. Density estimation. Problem: to estimate the probability density function of a continuous random variable

   • Kernel density estimation: a generalization of histograms

   • It is called "kernel" because the estimation of the density at point x is based on a kernel function that weights the observations around x. Typically, decreasing weights are attached to points further away from x.

2. Nonparametric regression:

   • A typical parametric regression is of the form:

$E(Y|X_1, X_2) = \beta_1 X_1 + \beta_2 X_2$

- A nonparametric regression has the form:

$E(Y|X_1, X_2) = m(X_1, X_2)$

Only assumption: $m(.)$ is a smooth function

3. Additive model (semiparametric regression):

$E(Y|X_1, X_2) = \alpha + m_1(X_1) + m_2(X_2)$

- Advantages of nonparametric methods: i) allow for estimation of more general functional forms; ii) useful when nonlinear effects are important

- Disadvantages: i) precision of estimates

Histograms

(Härdle et. al, 2004, Ch. 2)

- Histograms are nonparametric estimates of an unknown density function, $f(x)$. Procedure to build an histogram.

- Their crucial parameter is the binwidth h. A higher binwidth produces smoother estimates. It can be shown that the estimate is biased. The bias is positively related to h, while the variance of the estimate is negatively related to h.

  $h \uparrow \Rightarrow$ BIAS $\uparrow$. Insight: increasing h makes more and more difficult for the "bin" to approximate well the area under the smooth $f(x)$

  $h \uparrow \Rightarrow$ VARIANCE $\downarrow$. Insight: increasing h implies using more and more information to build the histrogram.

Problem: it is not possible to choose h in order to have a small bias and a small variance. Hence we need to find the "optimal" binwidth, which represents an optimal compromise.

- Two useful terms:

  • *oversmoothing*: obtained when h is large, reduction of variance but high bias;

  • *undersmoothing*: obtained when h is small, increase of variance but low bias;

Nonparametric density estimation

(Härdle et. al, 2004, Ch. 3)

- Problems with the histogram

1. each observation x in $\left[m_j - \frac{h}{2}, m_j + \frac{h}{2}\right]$ is estimated by the same value, $\widehat{f}_h(m_j)$.

2. $f(x)$ is estimated using the observations that fall in the interval containing x, and that receive the same weight in the estimation. That is, for $x \in B_j$,

   $$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} I\left(X_i \in B_j\right)$$

   where I is the indicator function

- Density estimation is based on the idea of generalizing an histogram.

- It is based on Kernel functions.

• The logic of using kernel functions is to estimate $f(x)$ using the observations that fall into an interval around x, which (typically) receive decreasing weight the further they are from x.

1. Kernel functions

   • Uniform kernel function: assigns the same weight to all observations in an interval of length $2h$ around observation x, $[x-h, x+h]$. That is, the estimate:

   $$\widehat{f}_h(x) = \frac{1}{2hn} \# \{X_i \in [x-h, x+h]\}$$

   can be obtained by means of a kernel function $K(u)$

   $$K(u) = \frac{1}{2} I(|u| \leq 1)$$

   where I is the indicator function and $u = (x - X_i)/h$. It assigns weight 1/2 to each observation X_i whose distance from x, the point where we want to estimate the density, is not bigger than h.

• A <u>Kernel function</u> (in general), assigns higher weights to observations in $[x-h, x+h]$ <u>closer to x</u>, e.g. Epanechnikov, Gaussian, etc.

• A kernel density estimation appears as a *sum of bumps*: at a given x, the value of $\widehat{f}_h(x)$ is found by vertically summing over the "bumps" (see Fig. 3.5 in Härdle et. al, 2004)

• In this case, we can write:

$$\widehat{f}_h(x) = \sum_{i=1}^{n} \frac{1}{nh} K\left(\frac{x - X_i}{h}\right) =$$
$$= \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

where $K_h(\cdot)$ is called "rescaled kernel function"

2. <u>Statistical properties of kernel density estimators</u>

• Same problems found for the histogram.

• Bias

$$Bias\left\{\widehat{f}_h(x)\right\} = E\left\{\widehat{f}_h(x)\right\} - f(x)$$

It can be shown that bias depends positively on h

- Variance

$$Var\left\{\widehat{f}_h(x)\right\} = Var\left\{\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\right\}$$

It can be shown that variance depends negatively on h

3. Choosing h

   (a) Define MSE (mean squared error)

   $$MSE\left\{\widehat{f}_h(x)\right\} = E\left[\left\{\widehat{f}_h(x) - f(x)\right\}^2\right]$$

   ...

   $$MSE = VAR\left\{\widehat{f}_h(x)\right\} + \left[Bias\left\{\widehat{f}_h(x)\right\}\right]^2$$

   Hence minimizing MSE may solve the trade-off, but the MSE-minimizing h depends on $f(x)$ and $f''(x)$, which are unknown.

(b) Define $MISE$ (mean integrated squared error). $MISE$ is preferable because it is a global measure of the error of the estimate.

$$MISE\left\{\hat{f}_h(x)\right\} =$$
$$= E\left[\int_{-\infty}^{\infty}\left\{\hat{f}_h(x) - f(x)\right\}^2 dx\right] =$$
$$= \int_{-\infty}^{\infty} MSE\left\{\hat{f}_h(x)\right\} dx$$

(c) Define $AMISE$ (an approximation of $MISE$) and obtain the formula for h_{opt}. The problem is that h_{opt} still depends on the unknown $f(x)$, in particular on its second derivative $f''(x)$.

(d) One possibility is a plug-in method suggested by Silverman, and consists in assuming that the unknown function is a Gaussian density function (whose variance is estimated by the sample variance). In this case h_{opt} has a simple formulation, and can be defined as a *rule-of-thumb* bandwidth.

Nonparametric regression

• Study the relation between two random variables: X (independent variable) and Y (dependent variable)

$y_i = m(x_i) + \epsilon_i, \; i = 1, .., n$

$E(Y|X = x) = m(x)$

• Kernel regression

• Consider the definition of the conditional expectation of Y given $X = x$:

$$E(Y|X = x) = \int y f(y|x) dy =$$
$$= \int y \frac{f(x, y)}{f_X(x)} dy = \frac{\int y f(x, y) dy}{f_X(x)} = m(x)$$

• To estimate $\widehat{m}(x)$, therefore, I need to estimate $f(x, y)$ and $f_X(x)$.

• Estimation of $f_X(x)$ is an instance of density estimation. Estimation of $f(x,y)$ requires the use of a "product kernel".

• One can obtain the Nadaraya-Watson estimator:

$$\hat{m}(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)}$$

where $K_h(x - X_i)$ is a kernel function

• This estimator can be defined as local mean estimator (see Bowman and Azzalini, 1997, p. 49). It can be obtained by solving the following problem:

$$min_\alpha \sum_{i=1}^{n} \{y_i - \alpha\}^2 K_h(x - X_i)$$

• The interpretation is that the various observations are replaced by a local mean, that is based on observations "close" to the point of estimation, where the weight that other observations have in determining the mean increases with their proximity to this point.

• It is possible to fit a *local linear regression*. In this case, the problem to solve is:

$$min_{\alpha,\beta} \sum_{i=1}^{n} \{y_i - \alpha - \beta(x - X_i)\}^2 K_h(x - X_i)$$

• It is also possible to fit a *local polynomial regression* (See Härdle at al., 2004, p. 94)

• It can be shown that increasing h produces smoother estimates. When $h \to 0$, then the estimates simply interpolates the points; when $h \to \infty$, the estimate is a constant function that assigns the sample mean of Y to each x.

• There are procedures to choose the optimal h

• With nonparametric methods it is possible to study (at least) two problems in growth empirics:

1. Parameter heterogeneity

2. Nonlinearity and multiple regimes

Some relevant papers

• Barro (1991), "Economic Growth in a Cross Section of Countries", QJE

• 98 countries observed in 1960-1985

• Method of analysis: cross-section regression

$\gamma_{i,T} = a + by_{i0} + \psi \mathbf{X}_i + \pi \mathbf{Z}_i + \epsilon_i$

• Dependent variable: average annual growth rate

• Explanatory variables:

1) initial GDP **(-)**

2) initial human capital (sec/prim) **(+)**

3) government consumption (should lower savings, no direct effect on productivity) **(-)**

4) indicators of political/social stability (should reduce investments) (revolutions/ assassinations) **(-)**

6) investment **(+)**

7) fertility **(-)**

8) index of political institutions (socialist/mixed) **SOC (-)**

9) continental dummies (Africa/Latin America) **(-)**

Interpretation of results

- A positive coefficient means that, holding fixed the other variables an increase in that variable has a positive marginal effect on the dependent variable

- Examples of the result on human capital: among countries with similar initial human capital (and the other variables), a higher initial income level is associated to lower growth. Among countries with similar initial income (and the other variables), a higher initial level of human capital is associated to higher growth.

- There is evidence of conditional convergence. The coefficient of initial income is negative, the set of controls includes the variables from the Solow model and other variables.

• Liu and Stengos (1999), "Non-linearities in Cross-Country Growth Regressions: a Semiparametric Approach" J. Appl. Econometrics

• 86 countries observed in 1960-1990

• Dependent variable: average annual growth rate

• Explanatory variables:

1) initial GDP

2) human capital

3) investment **(+)**

4) growth rate of population **(-)**

• Method of analysis: semiparametric regression

$$\gamma_i = \alpha + f_\beta(log(y_{i,0})) + \pi_n log(n_i + g + \delta) + \pi_K log s_{K,i} + f_{\pi_H}(log(s_{H,i})) + \epsilon_i$$

where $f_\beta(.)$ and $f_{\pi_H}(.)$ are arbitrary functions. The effect of $log(n_i + g + \delta)$ is estimated parametrically, the effects of $log y_{i,0}$ and $log s_{H,i}$ are estimated nonparametrically (this choice follows previous studies that highlighted the possible presence of thresholds in output and human capital, in particular Durlauf et al. 1995).

• This formulation follows Mankiw *et al.*, 1992:

$$\gamma_i = \alpha + \beta log(y_{i,0}) + \pi_n log(n_i + g + \delta) + \pi_K log(s_{K,i}) + \pi_H log(s_{H,i}) + \epsilon_i$$

• Liu and Stengos (1999) find that: (i) the effect of $log(y_{i,0})$ is negative only for incomes above \$1800; (ii) the effect of secondary school enrollment (empirical proxy for $log(s_{H,i})$) on growth is more pronounced when the variable is above 15% and weaker when the variable is above 75%. The relation may be linear for countries with a human capital level up to the intersection of the linear relation line and the confidence interval (show Figures).

• This approach useful to study nonlinearities, whose presence indicates parameter heterogeneity: "What do Thailand, the Dominican Republic, Zimbabwe, Greece and Bolivia have in common that merits their being put in the same regression analysis'?" (Harberger 1987, quoted in Durlauf *et al.*, 2004)

• Parameter heterogeneity may appear as a nonlinearity. A nonlinear effect simply means that the marginal effect of X on Y is different at different levels of X. If different countries have different levels of X, then the estimated coefficient on Y will differ.

Durlauf and Johnson (1995), “Multiple Regimes and Cross-Country Growth Behaviour”, J. Appl. Econometrics

• 96 countries observed in 1960-1985

• Dependent variable: average annual growth rate

• Explanatory variables (those in Mankiw et al., 1992):

1) initial GDP

2) Solow variables

3) human capital variables

• Method of analysis: 1) clustering of countries; 2) cross-section regression

• The aim of the paper is to determine “whether the data exhibit multiple regimes in the sense that sub-groups of countries identified by initial conditions obey distinct Solow-type regressions”

• First step: generate exogenous partitions of countries according to initial income and initial human capital and then run cross-section regression for each group. Result: the estimated coefficients are (very) different across the subgroups.

• Second step: check that the evidence of multiple regimes is not due to omitted variables (e.g. variables not included in the Solow model: country dummies, political variables, etc.). Run regressions in subgroups using additional variables. Results: adding controls does not change the previous result; countries with different initial conditions have different coefficients for the Solow variables

• Third step: generate an endogenous partition of countries in subgroups. In this case an algorithm (regression tree) is utilized. It produces four subgroups: 1) low income; 2) intermediate income/low literacy; 3) intermediate income/high literacy; 4) high income. Results show that the linear models estimated on the subgroups have very different coefficients, and probably obey different production functions

- On results: 1) the coefficient on initial income is negative and significant only for groups 1) and 3) implying convergence within them. 2) The human capital share is positive and significant only for groups 2) and 4). It may indicate the existence of technologies for which human capital is important (or simply that using only secondary school enrollment is inappropriate).

- Durlauf and Johnson (DJ) results vs the conditional convergence hypothesis (CCH): according to CCH countries with identical structural characteristics must converge to the same steady state independently of initial conditions. According to DJ, initial conditions determine structural characteristics, and therefore it cannot happen that one country may have some structural characteristics and any set of initial conditions.