Applied Economic Growth

Mario Lavezzi[1]

[1]University of Palermo
mario.lavezzi@unipa.it

International Doctoral Program in Economics
Scuola Superiore S. Anna
Pisa, June 2013

Table of Contents

Introduction

Issues

- Durlauf, Johnson and Temple (DJT)'s chapter in the Handbook of Economic Growth is entitled: "Growth econometrics". Why?

Introduction

Issues

- Durlauf, Johnson and Temple (DJT)'s chapter in the Handbook of Economic Growth is entitled: "Growth econometrics". Why?
- Because, besides many issues that are common in econometric analysis (measurment error, omitted variables, etc.), the empirical analysis of economic growth has specific problems.

Introduction

Issues

- Durlauf, Johnson and Temple (DJT)'s chapter in the Handbook of Economic Growth is entitled: "Growth econometrics". Why?
- Because, besides many issues that are common in econometric analysis (measurment error, omitted variables, etc.), the empirical analysis of economic growth has specific problems.
- The first is that there are few observations. "Large samples" include around 100 countries for approximately 50 years. Longer time series are available for only a subset of industrialized countries (Baumol, 1986, studies the period 1870-1979 for 16 countries)

Introduction

Issues

- This has implications not only for the precision of the estimates but also for more specific issues. For example, if we consider a policy issue such as: "should we export western democracy to poor countries?", then the number of observations we can count on is quite limited

Introduction

Issues

- This has implications not only for the precision of the estimates but also for more specific issues. For example, if we consider a policy issue such as: "should we export western democracy to poor countries?", then the number of observations we can count on is quite limited
- Moreover, documented growth experiences show remarkable heterogeneity in the cross-section and in time (especially for developing countries)

Introduction

Issues

- This has implications not only for the precision of the estimates but also for more specific issues. For example, if we consider a policy issue such as: "should we export western democracy to poor countries?", then the number of observations we can count on is quite limited
- Moreover, documented growth experiences show remarkable heterogeneity in the cross-section and in time (especially for developing countries)
- This represents a problem when we try to reconcile empirical evidence with theories, and when we try to use standard concepts for empirical analysis such as "trend" (Pritchett, 2000)

Introduction

Issues

- The main issues are:

Introduction

Issues

- The main issues are:
- Convergence: will currently poor countries catch up with the richest?

Introduction

Issues

- The main issues are:
- Convergence: will currently poor countries catch up with the richest?
- Model uncertainty: what significantly explains economic growth?

Introduction

Issues

- The main issues are:
- Convergence: will currently poor countries catch up with the richest?
- Model uncertainty: what significantly explains economic growth?
- Search for interesting patterns in data: parameter heterogeneity seems to be particularly important in the study of economic growth. This has led to the consideration of other statistical tools besides "standard" econometrics (for example: nonparametric methods, clustering algorithms)

Stylized Facts

Twin peaks/Polarization

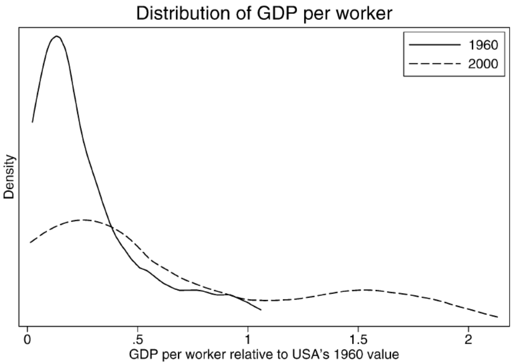


Figure 1. Cross-country density of output per worker.

- "Twin peaks": tendency for countries to "polarize" into "rich" and "poor", middle income group vanishing

Stylized Facts

Stability in relative positions

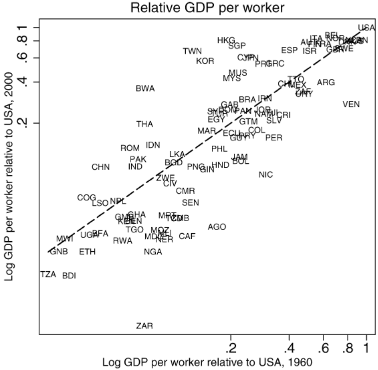


Figure 2. Output per worker: 1960 versus 2000.

- Most countries lie around the 45° line; "growth miracles" (above the 45° line); "growth disasters" (below the 45° line)

Stylized Facts

Growth miracles

Table 2
Fifteen growth miracles, 1960–2000

Country	Growth 1960–2000	Factor increase
Taiwan	6.25	11.3
Botswana	6.07	10.6
Hong Kong	5.67	9.09
Korea, Republic of	5.41	8.24
Singapore	5.09	7.29
Thailand	4.50	5.83
Cyprus	4.30	5.39
Japan	4.13	5.04
Ireland	4.10	5.00
China	3.99	4.77
Romania	3.91	4.63
Mauritius	3.88	4.58
Malaysia	3.82	4.48
Portugal	3.48	3.93
Indonesia	3.34	3.72

- Best 15 performers in the period: 1960-2000. Annual growth rate and ratio of GDP per worker in 2000 over its value in 1960. Mostly from East and Southeast Asia

Stylized Facts

Growth disasters

Table 3
Fifteen growth disasters, 1960–2000

Country	Growth 1960–2000	Ratio
Peru	0.00	1.00
Mauritania	−0.11	0.96
Senegal	−0.26	0.90
Chad	−0.43	0.84
Mozambique	−0.50	0.82
Madagascar	−0.60	0.79
Zambia	−0.61	0.78
Mali	−0.77	0.74
Venezuela	−0.88	0.70
Niger	−1.03	0.66
Nigeria	−1.21	0.62
Nicaragua	−1.30	0.59
Central African Republic	−1.56	0.53
Angola	−2.04	0.44
Congo, Democratic Rep.	−4.00	0.20

- Worst 15 performers in the period: 1960-2000. Annual growth rate and ratio of GDP per worker in 2000 over its value in 1960. Mostly from Sub-saharan Africa

Stylized Facts

Convergence?

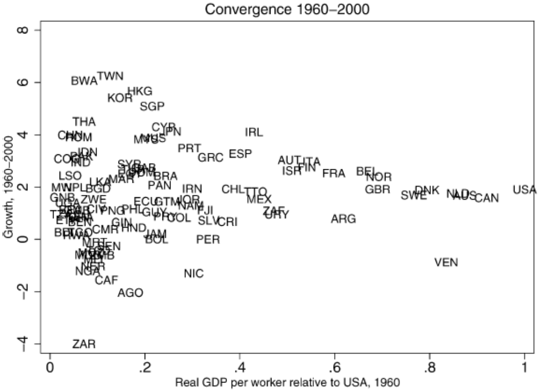


Figure 3. Growth versus initial income: 1960–2000.

- "Triangular shape" of the relation between growth rate and initial income. There is no negative relation

Stylized Facts

Diversity of growth

Table 7
Growth, 1960–2000, by country groups

Group	N	25th	Median	75th
Sub-Saharan Africa	36	−0.5	0.7	1.3
South and Central America	21	0.4	0.9	1.5
East and Southeast Asia	10	3.8	4.3	5.4
South Asia	7	1.9	2.2	2.9
Industrialized countries	19	1.7	2.4	3.0

Note: This table shows the 25th, 50th and 75th percentiles of the distribution of growth rates for various groups of countries.

- Diversity of growth experiences. Focus on regional variation

Stylized Facts

Growth volatility

Table 9
Volatility, 1960–2000, by regions

Group	*N*	25th	Median	75th
Sub-Saharan Africa	36	5.5	7.4	9.3
South and Central America	21	3.9	4.8	5.4
East and Southeast Asia	10	3.8	4.1	4.7
South Asia	7	3.0	3.3	5.2
Industrialized countries	19	2.3	2.9	3.5

Note: This table shows the 25th, 50th and 75th percentiles of the distribution of the standard deviation of annual growth rates, using data from the earliest available year until the latest available, between 1960 and 2000.

- Growth volatility is higher at low income levels. Focus on regional variation

Stylized Facts

Summary

- Growth disparities remain remarkable;

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong
- Variety of growth experiences:

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong
- Variety of growth experiences:
    - i growth miracles

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong
- Variety of growth experiences:
  - i growth miracles
  - ii growth disasters

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong
- Variety of growth experiences:
    - i growth miracles
    - ii growth disasters
    - iii regional disparities

Stylized Facts

Summary

- Growth disparities remain remarkable;
- Convergence in income does not seem to be strong
- Variety of growth experiences:
    - i growth miracles
    - ii growth disasters
    - iii regional disparities
    - iv differences in growth volatility

Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

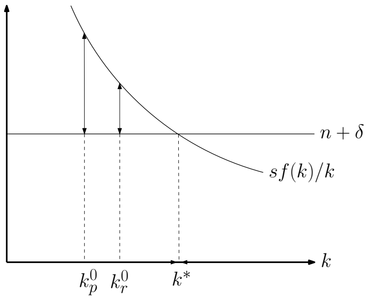
$$\frac{\dot{k}}{k} = \gamma_k = \frac{sf(k)}{k} - (n+\delta)$$

Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

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

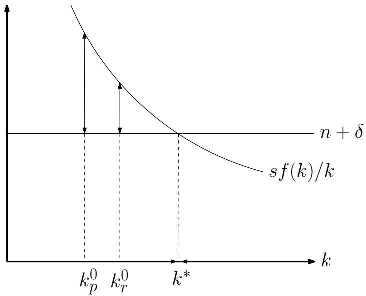


Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

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



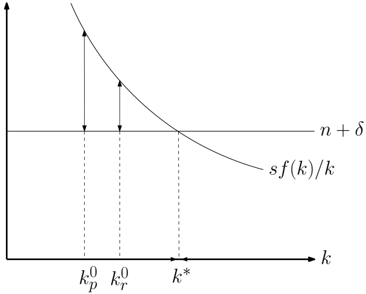
- Poor countries grow faster than rich countries (in the transition)

Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

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



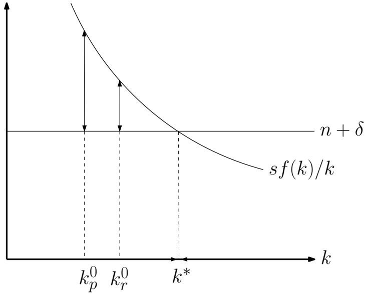
- Poor countries grow faster than rich countries (in the transition)
- Absolute convergence: per capita income of countries converge to one another in the long run, independently of their initial conditions (Galor, 1996)

Theoretical models of growth and convergence

The Solow model (no technological progress): absolute convergence

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

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



- Poor countries grow faster than rich countries (in the transition)
- Absolute convergence: per capita income of countries converge to one another in the long run, independently of their initial conditions (Galor, 1996)
- Transitory shocks on capital/income have no permanent effects

Theoretical models of growth and convergence

The Solow model (exogenous technological progress g): absolute convergence

$$\dot{k}^E = sf(k^E) - (n + g + \delta)k^E$$

Theoretical models of growth and convergence

The Solow model (exogenous technological progress g): absolute convergence

$$\dot{k^E} = sf(k^E) - (n + g + \delta)k^E$$

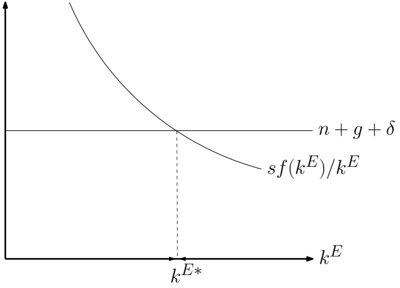
$$\frac{\dot{k^E}}{k^E} = \gamma_{k^E} = \frac{sf(k^E)}{k^E} - (n + g + \delta)$$

Theoretical models of growth and convergence

The Solow model (exogenous technological progress g): absolute convergence

$$\dot{k}^E = sf(k^E) - (n+g+\delta)k^E$$

$$\frac{\dot{k}^E}{k^E} = \gamma_{k^E} = \frac{sf(k^E)}{k^E} - (n+g+\delta)$$



Theoretical models of growth and convergence

Figure from Temple (1999)

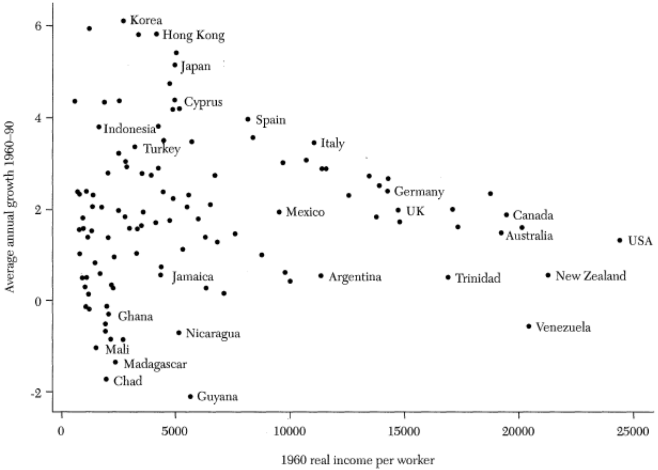


Figure 1. Growth and initial income, 1960–90

Theoretical models of growth and convergence

The Solow model: conditional convergence

▸ Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n+\delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n+\delta), s_r > s_p$$

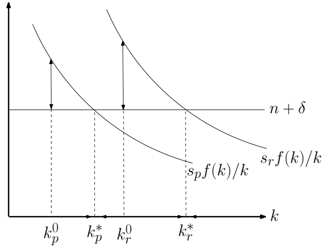
Theoretical models of growth and convergence

The Solow model: conditional convergence

Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n+\delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n+\delta), s_r > s_p$$



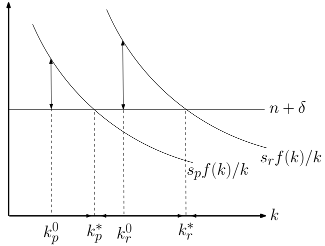
Theoretical models of growth and convergence

The Solow model: conditional convergence

Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n + \delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n + \delta), s_r > s_p$$



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

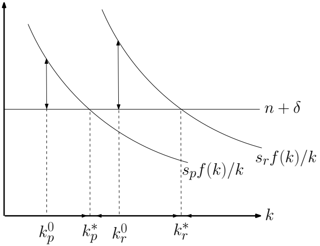
Theoretical models of growth and convergence

The Solow model: conditional convergence

▸ Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n+\delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n+\delta),\ s_r > s_p$$



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

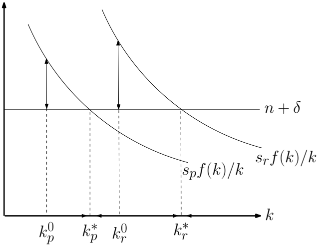
Theoretical models of growth and convergence

The Solow model: conditional convergence

Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n + \delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n + \delta), s_r > s_p$$



- Rich countries can grow faster than poor countries (in the transition)
- 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)
- Rejection of absolute convergence does not imply rejection of the Solow Model!

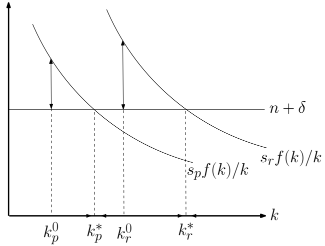
Theoretical models of growth and convergence

The Solow model: conditional convergence

▸ Cl. Conv.2

$$\frac{\dot{k}}{k} = \gamma_{k,r} = \frac{s_r f(k)}{k} - (n+\delta)$$

$$\frac{\dot{k}}{k} = \gamma_{k,p} = \frac{s_p f(k)}{k} - (n+\delta), s_r > s_p$$



- Rich countries can grow faster than poor countries (in the transition)
- 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)
- Rejection of absolute convergence does not imply rejection of the Solow Model!
- But ... observing persistent differences in income requires an explanation of persistent differences in structural parameters

Theoretical models of growth and convergence

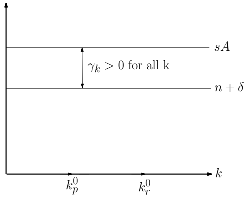
AK Model (endogenous growth)

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

Theoretical models of growth and convergence

AK Model (endogenous growth)

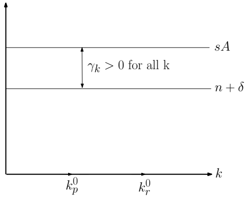
$$Y = AK$$
$$y = Ak$$
$$\frac{\dot{k}}{k} = \gamma_k = sA - (n + \delta)$$



Theoretical models of growth and convergence

AK Model (endogenous growth)

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

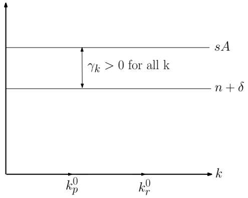


- No transitional dynamics

Theoretical models of growth and convergence

AK Model (endogenous growth)

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

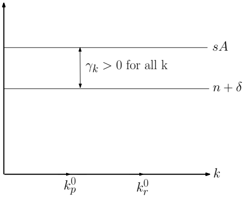


- No transitional dynamics
- No absolute convergence

Theoretical models of growth and convergence

AK Model (endogenous growth)

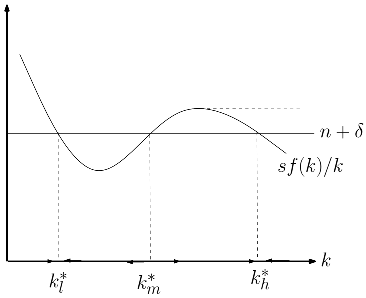
$$Y = AK$$
$$y = Ak$$
$$\frac{\dot{k}}{k} = \gamma_k = sA - (n + \delta)$$



- No transitional dynamics
- No absolute convergence
- No conditional convergence

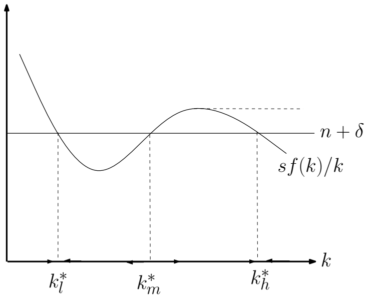
Theoretical models of growth and convergence

Multiple equilibria model



Theoretical models of growth and convergence

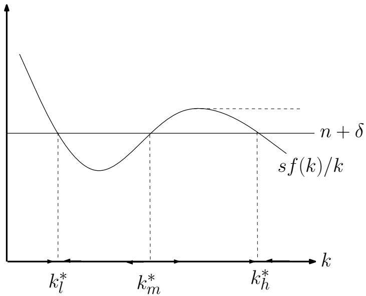
Multiple equilibria model



- 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)

Theoretical models of growth and convergence

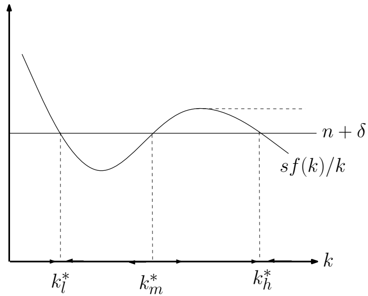
Multiple equilibria model



- 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)

Theoretical models of growth and convergence

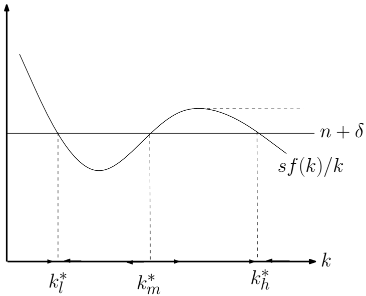
Multiple equilibria model



- 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 can have permanent effects

Theoretical models of growth and convergence

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 can have permanent effects
- The issue is the existence of an intermediate range of capital in which the relation growth rate/capital level is increasing ▸ Cond. Conv.

Theoretical models of growth and convergence

Multiple equilibria model

- Possible explanations:

Theoretical models of growth and convergence

Multiple equilibria model

- 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)

Theoretical models of growth and convergence

Multiple equilibria model

- 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.

Methods of empirical analysis

1. Growth regressions

Methods of empirical analysis

1. Growth regressions
2. Nonparametric methods

Methods of empirical analysis

1. Growth regressions
2. Nonparametric methods
    1. distribution dynamics

Methods of empirical analysis

1. Growth regressions
2. Nonparametric methods
    1. distribution dynamics
    2. nonparametric regressions

Growth regressions

Derivation of a growth regression

- Let $Y_{i,t}$ be the output, $L_{i,t}$ the labour force and $A_{i,t}$ the level of technology of country i at time t

Growth regressions

Derivation of a growth regression

- Let $Y_{i,t}$ be the output, $L_{i,t}$ the labour force and $A_{i,t}$ the level of technology of country i at time t
- Assume that $L_{i,t}$ and $A_{i,t}$ grow exogenously at rates n_i and g_i, that is: $L_{i,t} = L_{i,0}e^{n_i t}$; $A_{i,t} = A_{i,0}e^{g_i t}$

Growth regressions

Derivation of a growth regression

- Let $Y_{i,t}$ be the output, $L_{i,t}$ the labour force and $A_{i,t}$ the level of technology of country i at time t
- Assume that $L_{i,t}$ and $A_{i,t}$ grow exogenously at rates n_i and g_i, that is: $L_{i,t} = L_{i,0}e^{n_i t}$; $A_{i,t} = A_{i,0}e^{g_i t}$
- The generic one-sector growth model implies, to a first-order approximation, that:

$$\log(y^E_{i,t}) = (1 - e^{-\lambda_i t})\log(y^E_{i,\infty}) + e^{-\lambda_i t}\log(y^E_{i,0}), \qquad (1)$$

where $y^E_{i,\infty}$ is the steady-state value of $y^E_{i,t}$ (income in efficiency units), and the parameter λ_i measures the rate of convergence (Mankiw et al, 1992, p. 423, Barro and Sala-i-Martin, 2004, p. 58).

Growth regressions

Derivation of a growth regression

- Eq. (1) is expressed in terms $y^E_{i,t}$, which is unobservable. So rewrite Eq. (1) in terms of income per unit of labor, $y_{i,t}$:

$$\log(y_{i,t}) - g_i t - \log(A_{i,0}) = (1 - e^{-\lambda_i t})\log(y^E_{i,\infty}) + e^{-\lambda_i t}(\log(y_{i,0}) - \log(A_{i,0})) \quad (2)$$

Growth regressions

Derivation of a growth regression

- Eq. (1) is expressed in terms $y^E_{i,t}$, which is unobservable. So rewrite Eq. (1) in terms of income per unit of labor, $y_{i,t}$:

$$\log(y_{i,t}) - g_i t - \log(A_{i,0}) = (1 - e^{-\lambda_i t})\log(y^E_{i,\infty}) + e^{-\lambda_i t}(\log(y_{i,0}) - \log(A_{i,0})) \quad (2)$$

- From which:

$$\log(y_{i,t}) = g_i t + (1 - e^{-\lambda_i t})\log(y^E_{i,\infty}) + (1 - e^{-\lambda_i t})\log(A_{i,0}) + e^{-\lambda_i t}\log(y_{i,0}) \quad (3)$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Define $\gamma_i = t^{-1}[\log(y_{i,t}) - \log(y_{i,0})]$ and $\beta_i = -t^{-1}(1 - e^{-\lambda_i t})$, and subtract $\log(y_{i,0})$ from both sides of Eq. (3) then:

$$\gamma_i = g_i - \beta_i \log(y^E_{i,\infty}) - \beta_i \log(A_{i,0}) + \beta_i \log(y_{i,0}). \qquad (4)$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Define $\gamma_i = t^{-1}[\log(y_{i,t}) - \log(y_{i,0})]$ and $\beta_i = -t^{-1}(1 - e^{-\lambda_i t})$, and subtract $\log(y_{i,0})$ from both sides of Eq. (3) then:

$$\gamma_i = g_i - \beta_i \log(y^E_{i,\infty}) - \beta_i \log(A_{i,0}) + \beta_i \log(y_{i,0}). \quad (4)$$

- Finally, assuming parameter constancy across countries (i.e. $g_i = g$, $\lambda_i = \lambda \ \forall i$) we obtain:

$$\gamma_i = g - \beta \log(y^E_{i,\infty}) - \beta \log(A_{i,0}) + \beta \log(y_{i,0}). \quad (5)$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Define $\gamma_i = t^{-1}[\log(y_{i,t}) - \log(y_{i,0})]$ and $\beta_i = -t^{-1}(1 - e^{-\lambda_i t})$, and subtract $\log(y_{i,0})$ from both sides of Eq. (3) then:

$$\gamma_i = g_i - \beta_i \log(y^E_{i,\infty}) - \beta_i \log(A_{i,0}) + \beta_i \log(y_{i,0}). \qquad (4)$$

- Finally, assuming parameter constancy across countries (i.e. $g_i = g$, $\lambda_i = \lambda \ \forall i$) we obtain:

$$\gamma_i = g - \beta \log(y^E_{i,\infty}) - \beta \log(A_{i,0}) + \beta \log(y_{i,0}). \qquad (5)$$

- Eq. (5) is the starting point for a cross-country growth regression, after appending a random error term, that is:

$$\boxed{\gamma_i = g - \beta \log(y^E_{i,\infty}) - \beta \log(A_{i,0}) + \beta \log(y_{i,0}) + \nu_i.} \qquad (6)$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- In order to implement Eq. (6) it is necessary to empirically determine $y_{i,\infty}^{E}$ and $A_{i,0}$. Mankiw et al. (1992) show how to do this.

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- In order to implement Eq. (6) it is necessary to empirically determine $y^E_{i,\infty}$ and $A_{i,0}$. Mankiw et al. (1992) show how to do this.
- Consider a three-factor Cobb-Douglas production function for aggregate output:

$$Y_{i,t} = K^{\alpha}_{i,t} H^{\phi}_{i,t} (A_{i,t} L_{i,t})^{1-\alpha-\phi}, \tag{7}$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- In order to implement Eq. (6) it is necessary to empirically determine $y^E_{i,\infty}$ and $A_{i,0}$. Mankiw et al. (1992) show how to do this.
- Consider a three-factor Cobb-Douglas production function for aggregate output:

$$Y_{i,t} = K^{\alpha}_{i,t} H^{\phi}_{i,t} (A_{i,t} L_{i,t})^{1-\alpha-\phi}, \quad (7)$$

- where physical and human capital are accumulated following:

$$\dot{K}_{i,t} = s_{K,i} Y_{i,t} - \delta K_{i,t}; \quad (8)$$

$$\dot{H}_{i,t} = s_{H,i} Y_{i,t} - \delta H_{i,t}, \quad (9)$$

in which $s_{K,i}$ and $s_{H,i}$ are the saving rates for physical and human capital respectively.

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Equations (8)-(9) (with the parameter constancy assumptions) imply that economy converges to the steady-state value of output per effective worker:

$$y^E_{i,\infty} = \left(\frac{s^{\alpha}_{K,i} s^{\phi}_{H,i}}{(n_i + g + \delta)^{\alpha+\phi}} \right)^{\frac{1}{1-\alpha-\phi}} . \tag{10}$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Equations (8)-(9) (with the parameter constancy assumptions) imply that economy converges to the steady-state value of output per effective worker:

$$y^E_{i,\infty} = \left(\frac{s^{\alpha}_{K,i} s^{\phi}_{H,i}}{(n_i + g + \delta)^{\alpha+\phi}} \right)^{\frac{1}{1-\alpha-\phi}} . \quad (10)$$

- Substituting for $y^E_{i,\infty}$ in Eq. (6) we obtain:

$$\begin{aligned} \gamma_i &= g + \beta \log(y_{i,0}) + \beta \frac{\alpha+\phi}{1-\alpha-\phi} \log(n_i + g + \delta) - \\ &- \beta \frac{\alpha}{1-\alpha-\phi} \log(s_{K,i}) - \beta \frac{\phi}{1-\alpha-\phi} \log(s_{H,i}) - \\ &- \beta \log(A_{i,0}) + \nu_i \end{aligned} \quad (11)$$

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Mankiw et al. (1992) assume that $A_{i,0}$ is unobservable while $g + \delta$ is known

Growth regressions

Derivation of a growth regression: Mankiw et al. (1992)

- Mankiw et al. (1992) assume that $A_{i,0}$ is unobservable while $g+\delta$ is known
- In particular, $A_{i,0}$ should reflect not only technology, assumed to be constant across countries, but also country-specific differences which vary randomly (like climate, institutions, and so on), that is:

$$\log(A_{i,0}) = \log A + e_i, \tag{12}$$

where e_i is a country-specific shock independent of n_i, $s_{K,i}$ and $s_{H,i}$.

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- Then, Eq. (11) can be rewritten as:

$$\gamma_i = g - \beta \log(A) + \beta \log(y_{i,0}) + \beta \frac{\alpha + \phi}{1 - \alpha - \phi} \log(n_i + g + \delta) - \\ - \beta \frac{\alpha}{1 - \alpha - \phi} \log(s_{K,i}) - \beta \frac{\phi}{1 - \alpha - \phi} \log(s_{H,i}) + \epsilon_i, \qquad (13)$$

where $\epsilon_i = \nu_i - \beta e_i$.

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- Then, Eq. (11) can be rewritten as:

$$\gamma_i = g - \beta \log(A) + \beta \log(y_{i,0}) + \beta \frac{\alpha + \phi}{1 - \alpha - \phi} \log(n_i + g + \delta) - \\ - \beta \frac{\alpha}{1 - \alpha - \phi} \log(s_{K,i}) - \beta \frac{\phi}{1 - \alpha - \phi} \log(s_{H,i}) + \epsilon_i, \qquad (13)$$

  where $\epsilon_i = \nu_i - \beta e_i$.

- The canonical cross-country growth regression, may be seen as the "unconstrained version" of Eq. (13) where the cross-coefficient restrictions are ignored, that is:

$$\boxed{\gamma_i = \beta \log(y_{i,0}) + \psi X_i + \epsilon_i,} \qquad (14)$$

$$i = 1, \ldots, N$$

  where $X_i = (1, \log(n_i + g + \delta), \log(s_{K,i}), \log(s_{H,i}))$.

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- The variables $\log(y_{i,0})$ and those in X_i represent the growth determinants suggested by the Solow growth model (augmented in the case of Mankiw et al., 1992).

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- The variables $\log(y_{i,0})$ and those in X_i represent the growth determinants suggested by the Solow growth model (augmented in the case of Mankiw et al., 1992).
- However, many cross-country studies added additional control variables besides "Solow" variables. With respect to Mankiw et al. (1992), we can interpret this attempt as allowing for predictable heterogeneity in the steady-state growth rate g_i and the initial technology term $A_{i,0}$

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- In other words, $g_i - \beta log(A_{i,0})$ in Eq. (4) is not substituted with $g - \beta \log(A) - \beta e_i$, but with $g - \beta \log(A) + \pi Z_i - \beta e_i$. From this assumption we obtain:

$$\begin{aligned}\gamma_i &= g - \beta \log(A) + \beta \log(y_{i,0}) + \beta \frac{\alpha + \phi}{1 - \alpha - \phi} \log(n_i + g + \delta) - \\ &- \beta \frac{\alpha}{1 - \alpha - \phi} \log(s_{K,i}) - \beta \frac{\phi}{1 - \alpha - \phi} \log(s_{H,i}) + \boxed{\pi Z_i} + \\ &+ \epsilon_i, \end{aligned} \tag{15}$$

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- Notice, however, that regression in Eq. (15) does not identify whether controls Z_i are correlated with the steady-state level g_i or the initial level of technology $A_{i,0}$ (see DJT, p. 580, for a discussion).

Growth regressions

Derivation of growth regression in Mankiw et al. (1992)

- Notice, however, that regression in Eq. (15) does not identify whether controls Z_i are correlated with the steady-state level g_i or the initial level of technology $A_{i,0}$ (see DJT, p. 580, for a discussion).
- The baseline cross-country growth regressions found in many studies (sometimes defined: "Barro regressions"), may be seen as the "unconstrained version" of Eq. (15) where the cross-coefficient restrictions are ignored, that is:

$$\boxed{\gamma_i = \beta \log(y_{i,0}) + \psi X_i + \pi Z_i + \epsilon_i,} \quad i = 1, ..., N \tag{16}$$

where $\log(y_{i,0})$ and variables in X_i represent the "Solow" growth determinants, and Z_i represents other growth determinants

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

$$\gamma_i = a + by_{i0} + \epsilon_i$$

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

$$\gamma_i = a + by_{i0} + \epsilon_i$$

In this case: If $\hat{b} < 0 \rightarrow$ absolute convergence.

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

  $$\gamma_i = a + by_{i0} + \epsilon_i$$

  In this case: If $\hat{b} < 0 \rightarrow$ absolute convergence.

  2. Solow growth determinants only:

  $$\gamma_i = a + by_{i0} + \psi X_i + \epsilon_i$$

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

  $$\gamma_i = a + by_{i0} + \epsilon_i$$

  In this case: <u>If $\hat{b} < 0 \rightarrow$ absolute convergence</u>.
  2. Solow growth determinants only:

  $$\gamma_i = a + by_{i0} + \psi X_i + \epsilon_i$$

  X_i: Solow growth determinants.

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

     $$\gamma_i = a + by_{i0} + \epsilon_i$$

     In this case: If $\hat{b} < 0 \rightarrow$ absolute convergence.
  2. Solow growth determinants only:

     $$\gamma_i = a + by_{i0} + \psi X_i + \epsilon_i$$

     X_i: Solow growth determinants.If $\hat{b} < 0 \rightarrow$ conditional convergence.

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

     $$\gamma_i = a + by_{i0} + \epsilon_i$$

     In this case: If $\hat{b} < 0 \rightarrow$ absolute convergence.
  2. Solow growth determinants only:

     $$\gamma_i = a + by_{i0} + \psi X_i + \epsilon_i$$

     X_i: Solow growth determinants.If $\hat{b} < 0 \rightarrow$ conditional convergence.
  3. "Barro regression"

     $$\gamma_i = a + by_{i0} + \psi X_i + \pi Z_i + \epsilon_i$$

     Z_i: growth determinants not in the Solow model.

Growth regressions

Model specification

- To sum up: three possible types of equations can be estimated:
  1. No growth determinants:

     $$\gamma_i = a + by_{i0} + \epsilon_i$$

     In this case: If $\hat{b} < 0 \rightarrow$ absolute convergence.
  2. Solow growth determinants only:

     $$\gamma_i = a + by_{i0} + \psi X_i + \epsilon_i$$

     X_i: Solow growth determinants. If $\hat{b} < 0 \rightarrow$ conditional convergence.
  3. "Barro regression"

     $$\gamma_i = a + by_{i0} + \psi X_i + \pi Z_i + \epsilon_i$$

     Z_i: growth determinants not in the Solow model. If $\hat{b} < 0 \rightarrow$ conditional convergence.

Growth regressions

Some remarks

- General comment: X_i and 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 (initial) stock of physical and human capital, etc.

Growth regressions

Some remarks

- General comment: X_i and 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 (initial) 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 initial income $\rightarrow$ low level of democracy, low investment rates, etc.).

Growth regressions

Some remarks

- General comment: X_i and 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 (initial) 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 initial income $\rightarrow$ low level of democracy, low investment rates, etc.).
- Which variables should be included in Z_i?

Growth regressions

β-convergence

1. β-convergence:

Growth regressions

β-convergence

1. β-convergence:
   1. unconditional: economies with lower levels of per capita income tend to grow faster in per capita terms

Growth regressions

β-convergence

1. β-convergence:
   1. unconditional: economies with lower levels of per capita income tend to grow faster in per capita terms
   2. 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

Growth regressions

A remark on β-convergence

- Recall the definitions:

  1. β-convergence:
     $\hat{b} < 0$ without controls $\rightarrow$ unconditional β-convergence
     $\hat{b} < 0$ with controls $\rightarrow$ conditional β-convergence

Growth regressions

A remark on β-convergence

- Recall the definitions:
  1. β-convergence:
     $\hat{b} < 0$ without controls $\rightarrow$ unconditional β-convergence
     $\hat{b} < 0$ with controls $\rightarrow$ conditional β-convergence
  2. Bernard and Durlauf (1996) show that:

Growth regressions

A remark on β-convergence

- Recall the definitions:
  1. β-convergence:
     $\hat{b} < 0$ without controls $\rightarrow$ unconditional β-convergence
     $\hat{b} < 0$ with controls $\rightarrow$ conditional β-convergence
  2. Bernard and Durlauf (1996) show that:
     1. $\hat{b} < 0$ is not sufficient to conclude that there is β-convergence.

Growth regressions

A remark on β-convergence

- Recall the definitions:

  1. β-convergence:
     $\hat{b} < 0$ without controls $\rightarrow$ unconditional β-convergence
     $\hat{b} < 0$ with controls $\rightarrow$ conditional β-convergence
  2. Bernard and Durlauf (1996) show that:
     1. $\hat{b} < 0$ is not sufficient to conclude that there is β-convergence.
     2. The coefficient $\hat{b}$ is a weighted average. Some countries in the sample may follow the Solow model, some may not. The data can be generated from a model with multiple equilibria but the regression on the misspecified model can nonetheless return a negative $\hat{b}$. 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)

Nonparametric models

Introduction

- In general, nonparametric analysis allows for more flexibility in the estimated relationships

Nonparametric models

Introduction

- In general, nonparametric analysis allows for more flexibility in the estimated relationships
- This is obtained by carrying out the estimation *locally*, i.e. by using the information "near" the point where the estimation of a relationship should be made

Nonparametric models

Introduction

- In general, nonparametric analysis allows for more flexibility in the estimated relationships
- This is obtained by carrying out the estimation *locally*, i.e. by using the information “near” the point where the estimation of a relationship should be made
- Suggested readings: Bowman and Azzalini (1997) and Härdle et. al (2004)

Nonparametric models

Density estimation

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

Nonparametric models

Density estimation

- Density estimation. Problem: to estimate the probability density function of a continuous random variable Densities
- In particular, we will consider Kernel density estimation: a generalization of histograms

Nonparametric models

Density estimation

- Density estimation. Problem: to estimate the probability density function of a continuous random variable ▸ Densities
- In particular, we will consider 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

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

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

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$
- Procedure to build a histogram.
- we have n observations $X_1, X_2, ..., X_n$

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$
- Procedure to build a histogram.
- we have n observations $X_1, X_2, ..., X_n$
- select an origin x_0 on the real line and divide it into "bins" B_j of width h:

$$B_j = [x_0 + (j-1)h, x_0 + jh], \ j \in \mathbf{Z};$$

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$
- Procedure to build a histogram.
- we have n observations $X_1, X_2, ..., X_n$
- select an origin x_0 on the real line and divide it into "bins" B_j of width h:

$$B_j = [x_0 + (j-1)h, x_0 + jh],\ j \in \mathbf{Z};$$

- Count the observations in each B_j: n_j;

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$
- Procedure to build a histogram.
- we have n observations $X_1, X_2, ..., X_n$
- select an origin x_0 on the real line and divide it into "bins" B_j of width h:

$$B_j = [x_0 + (j-1)h, x_0 + jh], \; j \in \mathbf{Z};$$

- Count the observations in each B_j: n_j;
- for each bin divide the numbers n_j by the sample size n to obtain the relative frequencies $\frac{n_j}{n}$ then divide by h, so that the area under the histogram equals 1.

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Histograms are nonparametric estimates of an unknown density function, $f(x)$
- Procedure to build a histogram.
- we have n observations $X_1, X_2, ..., X_n$
- select an origin x_0 on the real line and divide it into "bins" B_j of width h:

$$B_j = [x_0 + (j-1)h, x_0 + jh], \; j \in \mathbf{Z};$$

- Count the observations in each B_j: n_j;
- for each bin divide the numbers n_j by the sample size n to obtain the relative frequencies $\frac{n_j}{n}$ then divide by h, so that the area under the histogram equals 1.
- Each bin of a histogram has height $f_j = \frac{n_j}{nh}$ and base h, so the area of bin B_j equals n_j/n

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Basically, a histogram assigns the same estimate $\hat{f}_h(x)$ to each x in a bin, based on the number of observations that fall in the bin containing x.

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Basically, a histogram assigns the same estimate $\hat{f}_h(x)$ to each x in a bin, based on the number of observations that fall in the bin containing x.
- formally, for $x \in B_j$,

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(X_i \in B_j) I(x \in B_j),$$

where I are indicator functions.

Nonparametric models

Histograms. Härdle et. al, 2004, Ch. 2

- Basically, a histogram assigns the same estimate $\hat{f}_h(x)$ to each x in a bin, based on the number of observations that fall in the bin containing x.
- formally, for $x \in B_j$,

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \sum_{j} I(X_i \in B_j) I(x \in B_j),$$

where I are indicator functions.

- Notice that, since observations in each B_j are counted, they receive the same weight in the estimation

Nonparametric models

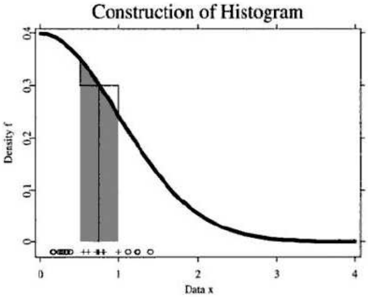
Histograms

- Denoting by m_j the centre of $B_j = [m_j - \frac{h}{2}, m_j + \frac{h}{2}]$, we can also say that a histogram assigns the estimation $\hat{f}_h(m_j)$ to every x in B_j

Nonparametric models

Histograms

- Denoting by m_j the centre of $B_j = [m_j - \frac{h}{2}, m_j + \frac{h}{2}]$, we can also say that a histogram assigns the estimation $\hat{f}_h(m_j)$ to every x in B_j

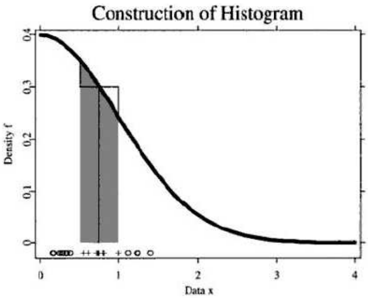


- Approximation of a density by histogram

Nonparametric models

Histograms

- Denoting by m_j the centre of $B_j = [m_j - \frac{h}{2}, m_j + \frac{h}{2}]$, we can also say that a histogram assigns the estimation $\hat{f}_h(m_j)$ to every x in B_j



- Approximation of a density by histogram

Nonparametric models

Histograms

- 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

Nonparametric models

Histograms

- 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
- $h \uparrow \Rightarrow \text{BIAS} \uparrow$. Intuition: increasing h makes more and more difficult for the "bin" to approximate well the area under the smooth function $f(x)$

Nonparametric models

Histograms

- 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 $\uparrow$. Intuition: increasing h makes more and more difficult for the "bin" to approximate well the area under the smooth function $f(x)$
- $h \uparrow \Rightarrow$ VARIANCE $\downarrow$. Intuition: increasing h implies using more and more information to build the histrogram.

Nonparametric models

Histograms

- 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.

Nonparametric models

Histograms

- 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;

Nonparametric models

Histograms

- 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 models

Histograms

- Problems with the histogram

Nonparametric models

Histograms

- 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, $\hat{f}_h(m_j)$.

Nonparametric models

Histograms

- 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, $\hat{f}_h(m_j)$.
  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 $x \in B_j$,

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

where I is the indicator function

Nonparametric models

From histograms to kernel density estimation

- In histograms: $f(x)$ is estimated by $1/nh$ times the number of observations into a small interval *containing* x

Nonparametric models

From histograms to kernel density estimation

- In histograms: $f(x)$ is estimated by $1/nh$ times the number of observations into a small interval *containing* x
- with kernels: $f(x)$ is estimated by $1/nh$ times the number of observations into a small interval *around* x

Nonparametric models

From histograms to kernel density estimation

- In histograms: $f(x)$ is estimated by $1/nh$ times the number of observations into a small interval *containing* x
- with kernels: $f(x)$ is estimated by $1/nh$ times the number of observations into a small interval *around* x
- In particular: kernels give more weigth in the estimation to points close to x (i.e. where the estimation should be carried out)

Nonparametric models

Kernel density estimation

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

Nonparametric models

Kernel density estimation

- 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:

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

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 $u = (x - X_i)/h$

Nonparametric models

Kernel density estimation

- 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:

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

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 $u = (x - X_i)/h$

Nonparametric models

Kernel density estimation

- 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.
- so we can write:

$$\hat{f}_h(x) = \frac{1}{hn}\sum_{1}^{n} K\left(\frac{x - X_i}{h}\right) \quad (17)$$

$$= \frac{1}{hn}\sum_{1}^{n} \frac{1}{2} I\left(\left|\frac{x - X_i}{h}\right| \leq 1\right)$$

Nonparametric models

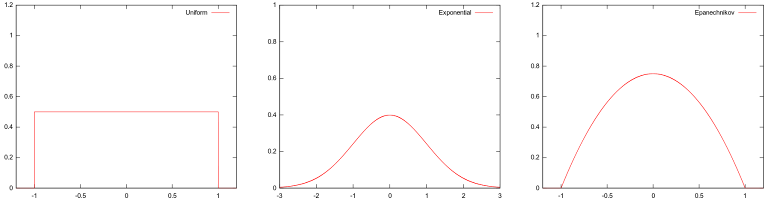
Kernel density estimation

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

Nonparametric models

Kernel density estimation

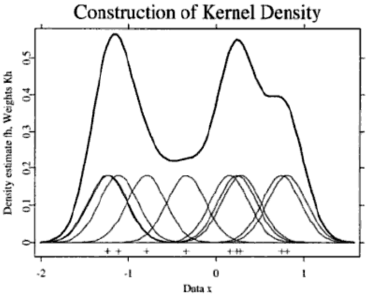
- A Kernel function (in general), assigns higher weights to observations in $[x-h, x+h]$ closer to x, e.g. Epanechnikov, Gaussian, etc.



- A kernel density estimation appears as a *sum of bumps*: at a given x, the value of $\hat{f}_h(x)$ is found by vertically summing over the "bumps"

Nonparametric models

Kernel density estimation



- In this case, we can write:

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

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

- Same problems found for the histogram

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

- Same problems found for the histogram
- Bias

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

- It can be shown that bias depends positively on h

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

- Same problems found for the histogram
- Bias

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

- It can be shown that bias depends positively on h
- Variance

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

- It can be shown that variance depends negatively on h

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

- Choosing h
- Define MSE (mean squared error)

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

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

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

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

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

$$MISE\left\{\hat{f}_h(x)\right\} = E\left[\int_{-\infty}^{\infty}\left\{\hat{f}_h(x) - f(x)\right\}^2 dx\right] = \int_{-\infty}^{\infty} MSE\left\{\hat{f}_h(x)\right\} dx \quad (18)$$

- 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 $f''(x)$.

Nonparametric models

Kernel density estimation: Statistical properties of kernel density estimators

- 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.

$$h^{opt} = \left(\frac{4}{3n} \right)^{\frac{1}{5}} \sigma$$

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

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

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

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

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

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

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

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

- Additive model (semiparametric regression):

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

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

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

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

- 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

Nonparametric models

Nonparametric regressions

- Parametric regression:

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

- Nonparametric regression:

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

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

- 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

Nonparametric models

Nonparametric regressions

Nonparametric models

Nonparametric regressions

- Study the relation between X (independent variable) and Y (dependent variable)

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

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

Nonparametric models

Nonparametric regressions

- Study the relation between 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

Nonparametric models

Nonparametric regressions

- Study the relation between 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

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

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x, y)$ and $f_X(x)$. One can obtain the Nadaraya-Watson estimator:

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x, y)$ and $f_X(x)$. One can obtain the Nadaraya-Watson estimator:

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

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x,y)$ and $f_X(x)$. 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

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x,y)$ and $f_X(x)$. 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:

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x, y)$ and $f_X(x)$. 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:

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

Nonparametric models

Nonparametric regressions

- To estimate $\hat{m}(x)$ I need to estimate $f(x,y)$ and $f_X(x)$. 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 values of the dependent variable 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.

Nonparametric models

Nonparametric regressions

Nonparametric models

Nonparametric regressions

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

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

Nonparametric models

Nonparametric regressions

- It is also 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)

Nonparametric models

Nonparametric regressions

- It is also 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)
- Notice that the Nadaraya-Watson estimator is a weighted sum of observations.

Distribution dynamics

Preliminary: σ-convergence

- 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:

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

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

Distribution dynamics

β-convergence vs σ-convergence

- β-convergence does not imply σ-convergence! (Barro and Sala-i-Martin, 2004, pp. 50-51)

Distribution dynamics

β-convergence vs σ-convergence

- β-convergence does not imply σ-convergence! (Barro and Sala-i-Martin, 2004, pp. 50-51)
- Consider:

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

$0 < b < 1$ implies absolute convergence, as the annual growth rate, $log(y_{i,t}/y_{i,t-1})$ is inversely related to $log(y_{i,t-1})$.

Distribution dynamics

β-convergence vs σ-convergence

- β-convergence does not imply σ-convergence! (Barro and Sala-i-Martin, 2004, pp. 50-51)
- Consider:

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

  $0 < b < 1$ implies absolute convergence, as the annual growth rate, $log(y_{i,t}/y_{i,t-1})$ is inversely related to $log(y_{i,t-1})$.
- Sample dispersion follows:

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

Distribution dynamics

β-convergence vs σ-convergence

- β-convergence does not imply σ-convergence! (Barro and Sala-i-Martin, 2004, pp. 50-51)
- Consider:

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

$0 < b < 1$ implies absolute convergence, as the annual growth rate, $log(y_{i,t}/y_{i,t-1})$ is inversely related to $log(y_{i,t-1})$.

- Sample 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 $\sigma_u^2 > 0$.

Distribution dynamics

β-convergence vs σ-convergence

- β-convergence does not imply σ-convergence! (Barro and Sala-i-Martin, 2004, pp. 50-51)
- Consider:

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

$0 < b < 1$ implies absolute convergence, as the annual growth rate, $log(y_{i,t}/y_{i,t-1})$ is inversely related to $log(y_{i,t-1})$.

- Sample 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 $\sigma_u^2 > 0$.

- The evolution of D_t follows:

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

D_t rises over time and converges to D^* if $D_0 < D^*$, even if $b > 0$.

Distribution dynamics

The distribution dynamics approach

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

Distribution dynamics

The distribution dynamics approach

- Starting point: to analyze the evolution of the whole income distribution
- Motivation: dissatisfaction with the standard approach based on cross-section regressions

Distribution dynamics

The distribution dynamics approach

- Starting point: to analyze the evolution of the whole income distribution
- Motivation: dissatisfaction with the standard approach based on cross-section regressions
- σ-convergence alone is informative on the dispersion of the distribution but non, for example, on within-distribution mobility

Distribution dynamics

The distribution dynamics approach

- Starting point: to analyze the evolution of the whole income distribution
- Motivation: dissatisfaction with the standard approach based on cross-section regressions
- σ-convergence alone is informative on the dispersion of the distribution but non, for example, on within-distribution mobility
- The distribution dynamics approach aims at highlighting convergence, divergence, intradistribution dynamics, catching up and falling behind

Distribution dynamics

The distribution dynamics approach

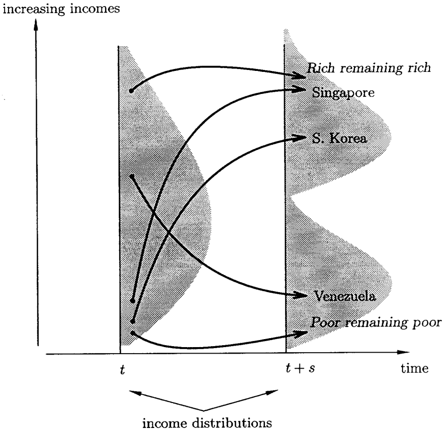


Fig. 1: Evolving cross-country income distributions Post-1960 experiences projected over 40 years for named countries are drawn to scale, relative to actual historical cross-country distributions.

Distribution dynamics

The distribution dynamics approach

- An example

Distribution dynamics

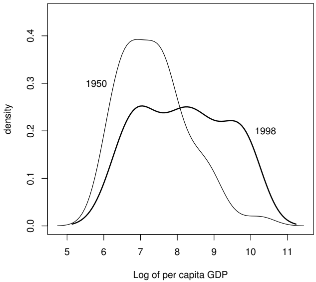
The distribution dynamics approach

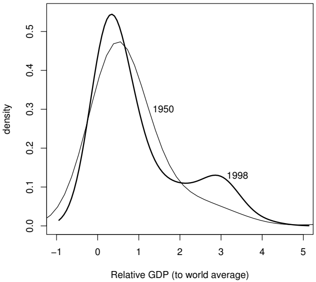
- An example
- Data from Maddison (2001), 122 countries, 1950-1998

Distribution dynamics

The distribution dynamics approach

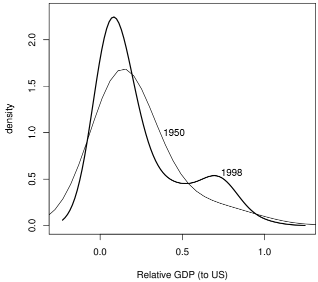
- An example
- Data from Maddison (2001), 122 countries, 1950-1998
- Estimates of density in 1950 and in 1998 ▸ IntroDensity

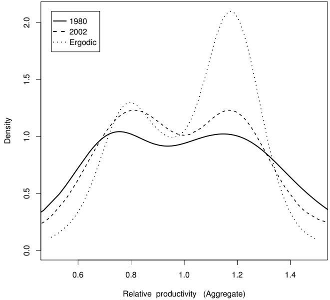




Distribution dynamics

The distribution dynamics approach

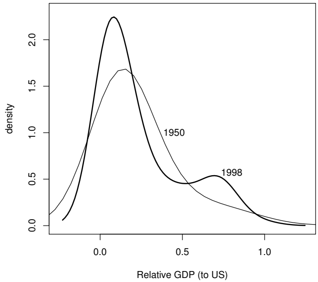


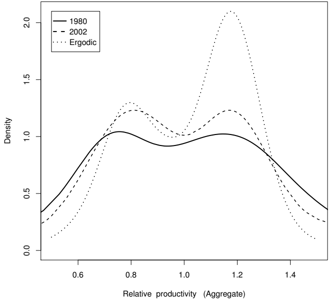


- Note the emergence of “twin peaks”

Distribution dynamics

The distribution dynamics approach

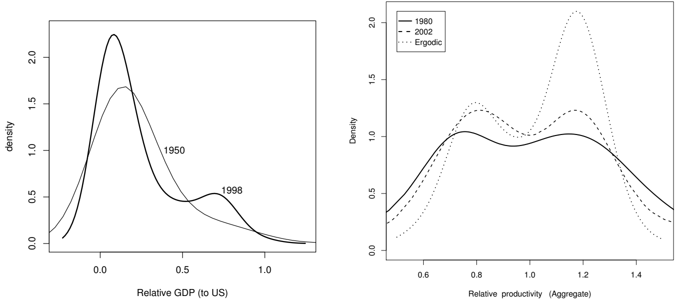




- Note the emergence of "twin peaks"
- This is considered as evidence of polarization: the distribution becomes thinner in the center and thicker in the tails

Distribution dynamics

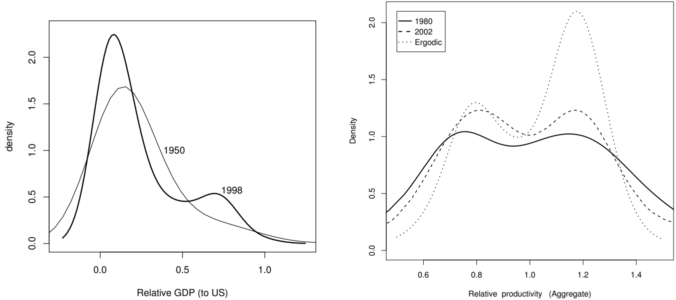
The distribution dynamics approach



- 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

Distribution dynamics

The distribution dynamics approach



- 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
- The distribution dynamics of European regions also displays two peaks (Fiaschi and Lavezzi, 2007)

Distribution dynamics

Markov Chains

Distribution dynamics

Markov Chains

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

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period
- Example: 3 income levels.

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period
- Example: 3 income levels.
    - State space of the process: $S = (1, 2, 3)$

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period
- Example: 3 income levels.
  - State space of the process: $S = (1, 2, 3)$
  - $q_t = [q_{1t}, q_{2t}, q_{3t}]$, $0 \leq q_{1t}, q_{2t}, q_{3t} \leq 1$,

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period
- Example: 3 income levels.
    - State space of the process: $S = (1, 2, 3)$
    - $q_t = [q_{1t}, q_{2t}, q_{3t}]$, $0 \leq q_{1t}, q_{2t}, q_{3t} \leq 1$,
    - $\sum_{i=1}^{3} q_{1t} = 1$

Distribution dynamics

Markov Chains

- 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 fraction of countries in that interval in a given period
- Example: 3 income levels.
    - State space of the process: $S = (1, 2, 3)$
    - $q_t = [q_{1t}, q_{2t}, q_{3t}]$, $0 \leq q_{1t}, q_{2t}, q_{3t} \leq 1$,
    - $\sum_{i=1}^{3} q_{1t} = 1$
- Dynamics is given by:

$$q_{t+1} = q_t \mathbf{P}$$

where $\mathbf{P}$ is a transition matrix

Distribution dynamics

Markov Chains

Distribution dynamics

Markov Chains

- Transition matrix:

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

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

Distribution dynamics

Markov Chains

- 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 **P** is a transition probability. It is a conditional probability:

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

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

Distribution dynamics

Markov Chains

- 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 **P** is a transition probability. It is a conditional probability:

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

  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:

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

Distribution dynamics

Markov Chains

- 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 **P** is a transition probability. It is a conditional probability:

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

  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:

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

- In this 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, i.e. as $\mathbf{P_t}$

Distribution dynamics

Markov Chains

Distribution dynamics

Markov Chains

- Long-run dynamics:

Distribution dynamics

Markov Chains

- Long-run dynamics:

$$q_1 = q_0 \mathbf{P}$$

$$q_2 = q_1 \mathbf{P} = q_0 \mathbf{P}^2$$

...

$$q_n = q_0 \mathbf{P}^n$$

- Under "regularity" conditions (see, e. g. Isaacson and Madsen, 1978):

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

and the process is ergodic.

Distribution dynamics

Markov Chains

- Long-run dynamics:

$$q_1 = q_0 \mathbf{P}$$

$$q_2 = q_1 \mathbf{P} = q_0 \mathbf{P}^2$$

...

$$q_n = q_0 \mathbf{P}^n$$

- Under "regularity" conditions (see, e. g. Isaacson and Madsen, 1978):

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

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

Distribution dynamics

Markov Chains

- A numerical example:

Distribution dynamics

Markov Chains

- A numerical example:
- A transition matrix such as:

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

Distribution dynamics

Markov Chains

- A numerical example:
- A transition matrix such as:

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

- has as invariant distribution:

$$\overline{q} = [0.41, 0.40, 0.18]$$

Distribution dynamics

Markov Chains

- A numerical example:
- A transition matrix such as:

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

- has as invariant distribution:

$$\overline{q} = [0.41, 0.40, 0.18]$$

- This is obtained by solving:

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

s.t. $\sum_{i=1}^{3} \overline{q}_i = 1$

Distribution dynamics

Quah (1993)

- From Quah (1993)

Distribution dynamics

Quah (1993)

- From Quah (1993)
- 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

Distribution dynamics

Quah (1993)

- From Quah (1993)
- 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

Distribution dynamics

Quah (1993)

- From Quah (1993)
- 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)

Distribution dynamics

Quah (1993)

- From Quah (1993)
- 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.

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

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

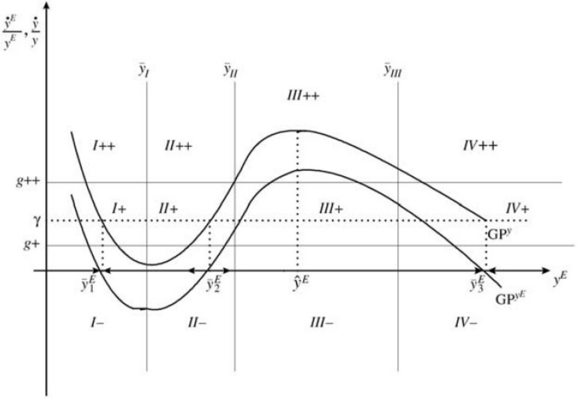
Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- 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.

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process



Distribution dynamics

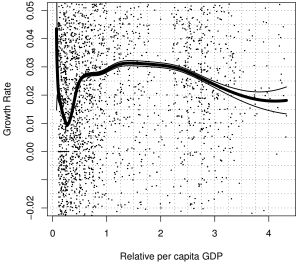
Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

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

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

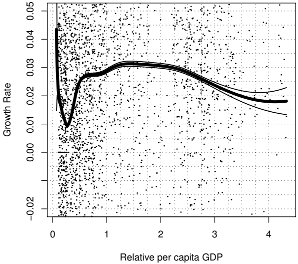
- First step: run a nonparametric regression of growth rates against income levels



Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

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



- Second step: study the distribution dynamics

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- Definition of the state space

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- 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++

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

Table: 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

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- Distribution dynamics and ergodic distribution

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- 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

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- 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

Distribution dynamics

Fiaschi and Lavezzi (2003): distribution dynamics and the shape of the growth process

- 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

Continuous state space

- See Johnson, 2005, and Johnson's webpage

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- Problem: discretization of state space may distort the underlying dynamics, especially in the long run

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- Problem: discretization of state space may distort the underlying dynamics, especially in the long run
- Possible solution: avoid discretization

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- 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:

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- 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

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- 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 $t+\tau$

Distribution dynamics

Continuous state space

- See Johnson, 2005, and Johnson's webpage
- 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 $t+\tau$

Distribution dynamics

Continuous state space

Distribution dynamics

Continuous state space

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

Distribution dynamics

Continuous state space

- 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

Distribution dynamics

Continuous state space

- 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 analog of a transition matrix. It maps the distribution of time t into the distribution at time $t + \tau$

Distribution dynamics

Continuous state space

- 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 analog of a transition matrix. It maps the distribution of time t into the distribution at time $t + \tau$
- it is denoted as stochastic kernel

Distribution dynamics

Continuous state space

- 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 analog of a transition matrix. It maps the distribution of time t into the distribution at time $t + \tau$
- it is denoted as stochastic kernel
- If the ergodic distribution implied by $g_\tau(x|z)$ exists, $f_\infty(x)$, it satisfies:

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

Distribution dynamics

Continuous state space

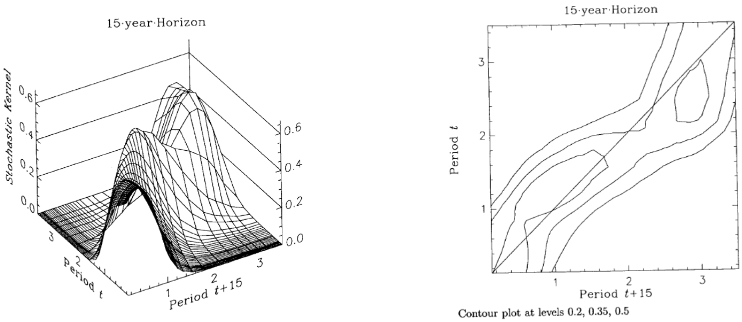


Figure: from Durlauf and Quah (1999)

Distribution dynamics

On the determinants of distribution dynamics

- Once we observe polarization, we can ask ourselves what are the explanatory variables

Distribution dynamics

On the determinants of distribution dynamics

- Once we observe polarization, we can ask ourselves what are the explanatory variables
- Attempts to tackle this issue in the literature are: Quah (1996 and 1997) in a "classic" distribution dynamics framework

Distribution dynamics

On the determinants of distribution dynamics

- Once we observe polarization, we can ask ourselves what are the explanatory variables
- Attempts to tackle this issue in the literature are: Quah (1996 and 1997) in a "classic" distribution dynamics framework
- Quah (1996) computes *conditioned stochastic kernels* based on residuals from two-sided regressions of labor productivity on human capital, physical capital, and country dummies

Distribution dynamics

On the determinants of distribution dynamics

- Once we observe polarization, we can ask ourselves what are the explanatory variables
- Attempts to tackle this issue in the literature are: Quah (1996 and 1997) in a "classic" distribution dynamics framework
- Quah (1996) computes *conditioned stochastic kernels* based on residuals from two-sided regressions of labor productivity on human capital, physical capital, and country dummies
- Quah (1997) defines conditioned stochastic kernels as operators mapping *unconditioned* income levels into *conditioned* income levels, that is incomes normalized: "on the basis of incomes relative to one's neighbours appropriately weighted" (p. 47) where weights are calculated with respect to one of the variables affecting the income dynamics
- Beaudry et al (2005) and Fiaschi et al (2013), instead, utilize

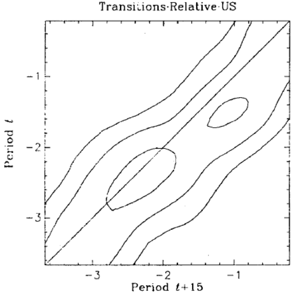
Distribution dynamics

On the determinants of distribution dynamics

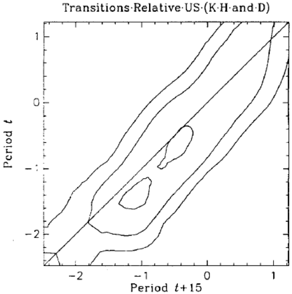
- Once we observe polarization, we can ask ourselves what are the explanatory variables
- Attempts to tackle this issue in the literature are: Quah (1996 and 1997) in a "classic" distribution dynamics framework
- Quah (1996) computes *conditioned stochastic kernels* based on residuals from two-sided regressions of labor productivity on human capital, physical capital, and country dummies
- Quah (1997) defines conditioned stochastic kernels as operators mapping *unconditioned* income levels into *conditioned* income levels, that is incomes normalized: "on the basis of incomes relative to one's neighbours appropriately weighted" (p. 47) where weights are calculated with respect to one of the variables affecting the income dynamics
- Beaudry et al (2005) and Fiaschi et al (2013), instead, utilize

Distribution dynamics

On the determinants of distribution dynamics: Quah (1996)



(b)

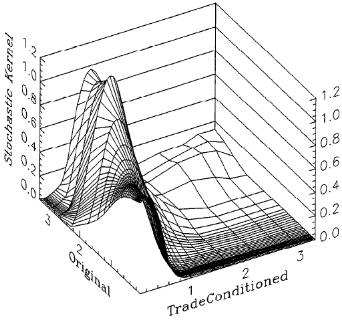


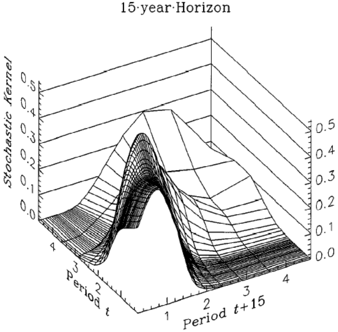
(b)

- "Comparing unconditional and conditional kernels (Figures 5 and 6) one sees that fine details differ, but the global dynamics of the distribution remain roughly unchanged. There are the same polarization, persistence, and immobility features in both. While the conditioning variables do affect the behavior of productivities in each country, they do not affect the dynamics of the entire distribution" (p. 114)

Distribution dynamics

On the determinants of distribution dynamics: Quah (1997)

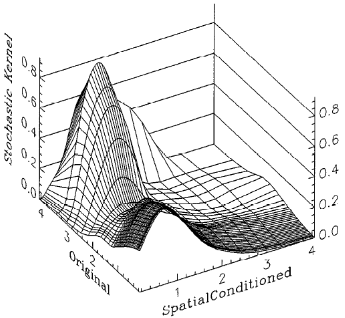


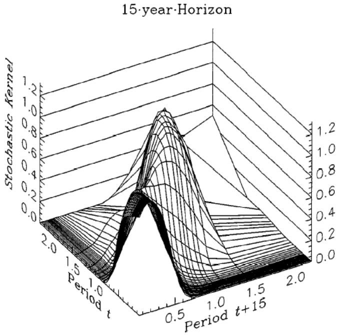


- "The most prominent change comparing stochastic kernels in Figure 7.1 (conditioned) and Figure 5.1 (unconditioned) is the counterclockwise shift in mass to parallel the Original axis. Put differently, spatial factors account for a large part of the distribution of incomes across countries: rich economies are typically close tointeract more withother rich ones; similarly poor economies are typically close to other poor ones ... — S displayed in Figure 7.2 and Figure 7.3 no longer show emerging twin-peaks features. In summary, it appears that the polarization earlier identified in the unconditional distribution-dynamics

Distribution dynamics

On the determinants of distribution dynamics: Quah (1997)





- "Here, the counterclockwise twist in the kernel towards the vertical is even more pronounced than in Figure 7.1: rich countries trade mostly with other rich ones; and, interestingly, the very poorest countries, mostly with rich ones again."
- "For trade, however, that increase in convergence dynamics is most obvious only for middle-income countries"

Distribution dynamics

On the determinants of distribution dynamics: Beaudry et al. (1995)

- Beaudry et al. (2005) consider two periods: 1960-1978 and 1978-1998: the second is characterized by a tendency for polarization

Distribution dynamics

On the determinants of distribution dynamics: Beaudry et al. (1995)

- Beaudry et al. (2005) consider two periods: 1960-1978 and 1978-1998: the second is characterized by a tendency for polarization
- They define an actual and a counterfactual growth rate:

$$g_i^{78-98} = \beta_0^{78-98} + \beta_y^{78-98} y_i^{78} + X_i^{78-98} \beta_x^{78-98} + \epsilon_i^{78-98}$$

$$g_i^{\beta_x} = \beta_0^{78-98} + \beta_y^{78-98} y_i^{78} + X_i^{78-98} \boxed{\beta_x^{60-78}} + \epsilon_i^{78-98}$$

Distribution dynamics

On the determinants of distribution dynamics: Beaudry et al. (1995)

- Beaudry et al. (2005) consider two periods: 1960-1978 and 1978-1998: the second is characterized by a tendency for polarization
- They define an actual and a counterfactual growth rate:

$$g_i^{78-98} = \beta_0^{78-98} + \beta_y^{78-98} y_i^{78} + X_i^{78-98} \beta_x^{78-98} + \epsilon_i^{78-98}$$

$$g_i^{\beta_x} = \beta_0^{78-98} + \beta_y^{78-98} y_i^{78} + X_i^{78-98} \boxed{\beta_x^{60-78}} + \epsilon_i^{78-98}$$

- and the related counterfactual income in 1998 (given $y_i^{98} = y_i^{78} + 20 g_i^{78-98}$:

$$y^{\beta_x} = y_i^{78} + 20 g_i^{\beta_x} = y_i^{98} + 20 X_i^{78-98} (\beta_x^{60-78} - \beta_x^{78-98})$$

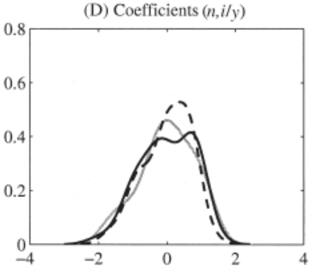
Distribution dynamics

On the determinants of distribution dynamics: Beaudry et al. (1995)

TABLE 3.—GROWTH REGRESS

	OLS		WLS	
	60–78	78–98	60–78	78–98
Const.	0.185	0.208	0.161	0.213
	(0.025)	(0.026)	(0.023)	(0.021)
y_0	−0.013	−0.014	−0.010	−0.015
	(0.002)	(0.002)	(0.002)	(0.002)
n	−0.137	−0.606	−0.118	−0.623
	(0.164)	(0.163)	(0.173)	(0.161)
i/y	0.016	0.027	0.018	0.025
	(0.003)	(0.004)	(0.003)	(0.004)
R^2	0.36	0.52	0.43	0.62
$\mathcal{Q}$(total)	14.642	[0.002]	8.590	[0.035]
$\mathcal{Q}(y_0)$	0.033	[0.855]	2.692	[0.101]
$\mathcal{Q}(n, i/y)$	12.021	[0.002]	8.212	[0.016]

Note: Standard errors in parentheses, *p*-values in brackets. WLS: Weights correspond to the log of the country's populat
IV2: The variable used for instrumenting i/y is the average of (c/y) over each subperiod. 75 observations.



- Their result is that, had the coefficients on physical capital and labor force remained the same, the distribution in 1998 would have been single-peaked

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- Labour productivity: $y_i(T) = y_i(0)e^{g_i T}$.
- Growth rate g_i:

$$g_i = m(\mathbf{X}_i) + \upsilon_i = \alpha + \sum_{k=1}^{K} \mu_k(X_{i,k}) + \upsilon_i$$

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- Labour productivity: $y_i(T) = y_i(0)e^{g_i T}$.
- Growth rate g_i:

$$g_i = m(\mathbf{X}_i) + \upsilon_i = \alpha + \sum_{k=1}^{K} \mu_k(X_{i,k}) + \upsilon_i$$

- i.e. we use a semiparametric specification

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- Define $\mathbf{X}_{i,\underline{k}} = (X_{i,1}, \ldots, X_{i,(k-1)}, X_{i,(k+1)}, \ldots, X_{i,K})$. Substituting:

$$\begin{aligned} y_i(T) &= y_i(0)e^{[\alpha+\mu_k(X_{i,k})+\sum_{j\neq k}\mu_j(X_{i,j})+\upsilon_i]T} = \\ &= \underbrace{y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}}_{y_{i,\underline{k}}(T)} \underbrace{e^{\mu_k(X_{i,k})T}}_{e^{g^M_{i,k}T}} \underbrace{e^{\upsilon_i T}}_{e^{g^R_i T}}, \end{aligned} \tag{19}$$

- where $y_{i,\underline{k}}(T) = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}$ is the level of productivity in period T obtained by "factoring out" the effect of $X_{i,k}$;

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- Define $\mathbf{X}_{i,\underline{k}} = (X_{i,1}, \dots, X_{i,(k-1)}, X_{i,(k+1)}, \dots, X_{i,K})$. Substituting:

$$\begin{aligned} y_i(T) &= y_i(0)e^{[\alpha+\mu_k(X_{i,k})+\sum_{j\neq k}\mu_j(X_{i,j})+\upsilon_i]T} = \\ &= \underbrace{y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}}_{y_{i,\underline{k}}(T)} \underbrace{e^{\mu_k(X_{i,k})T}}_{e^{g^M_{i,k}T}} \underbrace{e^{\upsilon_i T}}_{e^{g^R_i T}}, \end{aligned} \tag{19}$$

- where $y_{i,\underline{k}}(T) = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}$ is the level of productivity in period T obtained by "factoring out" the effect of $X_{i,k}$;
- $g^M_{i,k} = \mu_k(X_{i,k})$ is the part of the annual growth rate of y_i explained by $X_{i,k}$, capturing the "marginal" effect of $X_{i,k}$ on g_i;

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- Define $\mathbf{X}_{i,\underline{k}} = (X_{i,1}, ..., X_{i,(k-1)}, X_{i,(k+1)}, ..., X_{i,K})$. Substituting:

$$\begin{aligned} y_i(T) &= y_i(0)e^{[\alpha+\mu_k(X_{i,k})+\sum_{j\neq k}\mu_j(X_{i,j})+\upsilon_i]T} = \\ &= \underbrace{y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}}_{y_{i,\underline{k}}(T)} \underbrace{e^{\mu_k(X_{i,k})T}}_{e^{g^M_{i,k}T}} \underbrace{e^{\upsilon_i T}}_{e^{g^R_i T}}, \end{aligned} \qquad (19)$$

- where $y_{i,\underline{k}}(T) = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}$ is the level of productivity in period T obtained by "factoring out" the effect of $X_{i,k}$;
- $g^M_{i,k} = \mu_k(X_{i,k})$ is the part of the annual growth rate of y_i explained by $X_{i,k}$, capturing the "marginal" effect of $X_{i,k}$ on g_i;
- $g^R_i = \upsilon_i$ is the annual "residual growth", not explained by the variables in $\mathbf{X}_i$

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The counterfactual productivity at time T, $y_{i,k}^{CF}(T)$ is defined as:

$$y_{i,k}^{CF}(T) \equiv y_i(0)e^{g_{i,k}^{CF}T} = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})+\mu_k(\bar{X}_k)+\upsilon_i]T}$$

where $\bar{X}_k = N^{-1}\sum_{i=1}^{N} X_{i,k}$.

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The counterfactual productivity at time T, $y_{i,k}^{CF}(T)$ is defined as:

$$y_{i,k}^{CF}(T) \equiv y_i(0)e^{g_{i,k}^{CF}T} = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})+\mu_k(\bar{X}_k)+\upsilon_i]T}$$

where $\bar{X}_k = N^{-1}\sum_{i=1}^{N} X_{i,k}$.

- Counterfactual productivities are those that would have been obtained had all the countries had the same value of the variable X (supposed equal to its mean)

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The counterfactual productivity at time T, $y_{i,k}^{CF}(T)$ is defined as:

$$y_{i,k}^{CF}(T) \equiv y_i(0)e^{g_{i,k}^{CF}T} = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})+\mu_k(\bar{X}_k)+\upsilon_i]T}$$

where $\bar{X}_k = N^{-1}\sum_{i=1}^{N} X_{i,k}$.

- Counterfactual productivities are those that would have been obtained had all the countries had the same value of the variable X (supposed equal to its mean)
- Counterfactual productivities are the bases to compute *counterfactual stochastic kernels*.

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The actual stochastic kernel $\phi(\cdot)$ maps the distribution of (relative) productivity in period 0 into the distribution of (relative) productivity in period T.

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The actual stochastic kernel $\phi(\cdot)$ maps the distribution of (relative) productivity in period 0 into the distribution of (relative) productivity in period T.
- The counterfactual stochastic kernel $\phi^{CF}(\cdot)$, instead, maps the distribution of (relative) productivity in period 0, into the distribution of counterfactual relative productivities in period T.

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)

- The actual stochastic kernel $\phi(\cdot)$ maps the distribution of (relative) productivity in period 0 into the distribution of (relative) productivity in period T.
- The counterfactual stochastic kernel $\phi^{CF}(\cdot)$, instead, maps the distribution of (relative) productivity in period 0, into the distribution of counterfactual relative productivities in period T.
- Therefore, the counterfactual stochastic kernel shows, for every initial productivity level, the probability distribution over productivity levels at time T had the cross-country heterogeneity in the variable k been absent.

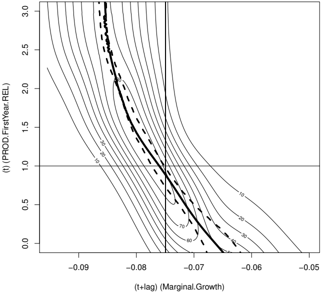
Distribution dynamics

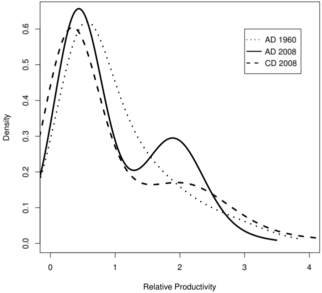
On the determinants of distribution dynamics: Fiaschi et al. (2013)

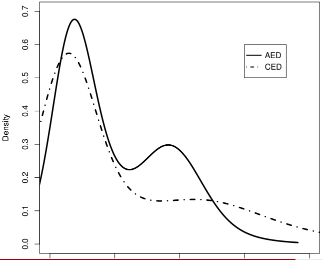
- The actual stochastic kernel $\phi(\cdot)$ maps the distribution of (relative) productivity in period 0 into the distribution of (relative) productivity in period T.
- The counterfactual stochastic kernel $\phi^{CF}(\cdot)$, instead, maps the distribution of (relative) productivity in period 0, into the distribution of counterfactual relative productivities in period T.
- Therefore, the counterfactual stochastic kernel shows, for every initial productivity level, the probability distribution over productivity levels at time T had the cross-country heterogeneity in the variable k been absent.
- This implies that the possible differences with respect to the probability distribution based on the actual stochastic kernel depends on the *k-th* variable, in particular on its distribution across countries.

Distribution dynamics

On the determinants of distribution dynamics: Fiaschi et al. (2013)







	AD 1960	AD 2008	CD 2008
Gini	0.42	0.42	0.52
s.e.	(0.029)	(0.026)	(0.030)
$BIPOL_{NI}$	*NA*	0.69	0.62
s.e.	(-)	(0.087)	(0.329)
Unimodality Test			
p-value	0.271	0.009	0.092
		AED	CED
Gini		0.46	0.57
s.e	.	(0.335)	(0.457)
$BIPOL_{NI}$		0.70	0.69
s.e	.	(0.102)	(0.560)
Unimodality Test			
p-value		0.006	0.092

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

$$\gamma_{i,T} = a + by_{i0} + \psi X_i + \pi Z_i + \epsilon_i$$

- Dependent variable: average annual growth rate
- Explanatory variables (sign of the estimated coefficient in parenthesis):
  1. initial GDP **(-)**
  2. initial human capital (sec/prim) **(+)**
  3. government consumption **(-)**
  4. indicators of political/social stability (should negatively affect property rights and reduce investments) (revolutions/ assassinations) **(-)**
  5. investment **(+)**
  6. population growth **(-)**
  7. index of political institutions (socialist/mixed) **SOC** (-)
  8. continental dummies (Africa/Latin America) **(-)**

Some relevant papers

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

- Interpretation of results

Some relevant papers

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

- 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

Some relevant papers

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

- 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.

Some relevant papers

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

- 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.

Some relevant papers

Mankiw, Romer and Weil (1992), "A Contribution to the Empirics of Economic Growth', QJE

- The estimated equation is:

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

TABLE V
TESTS FOR CONDITIONAL CONVERGENCE

Dependent variable: log difference GDP per working-age person 1960–1985

Sample:	Non-oil	Intermediate	OECD
Observations:	98	75	22
CONSTANT	3.04	3.69	2.81
	(0.83)	(0.91)	(1.19)
ln(Y60)	−0.289	−0.366	−0.398
	(0.062)	(0.067)	(0.070)
ln(I/GDP)	0.524	0.538	0.335
	(0.087)	(0.102)	(0.174)
$\ln(n+g+\delta)$	−0.505	−0.551	−0.844
	(0.288)	(0.288)	(0.334)
ln(SCHOOL)	0.233	0.271	0.223
	(0.060)	(0.081)	(0.144)
$\bar{R}^2$	0.46	0.43	0.65
s.e.e.	0.33	0.30	0.15
Implied λ	0.0137	0.0182	0.0203
	(0.0019)	(0.0020)	(0.0020)

Note. Standard errors are in parentheses. Y60 is GDP per working-age person in 1960. The investment and population growth rates are averages for the period 1960–1985. $(g+\delta)$ is assumed to be 0.05. SCHOOL is the average percentage of the working-age population in secondary school for the period 1960–1985.

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

- 111 countries observed in 1985/1992

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

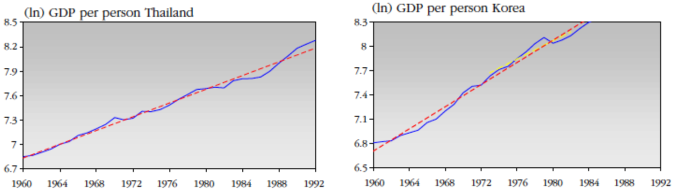
- 111 countries observed in 1985/1992
- Aim: emphasizing diversity in growth experiences

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

- 111 countries observed in 1985/1992
- Aim: emphasizing diversity in growth experiences

Figure 4a. Steep hills

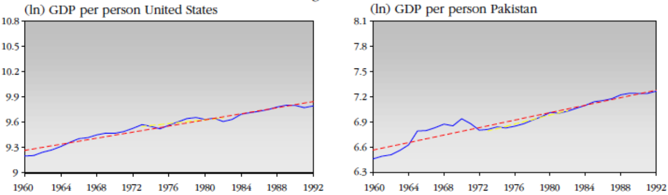


- Steep Hills: "These 11 countries had growth rates higher than 3 percent in both periods ... In these countries the trend is everything"

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

Figure 4b. Hills

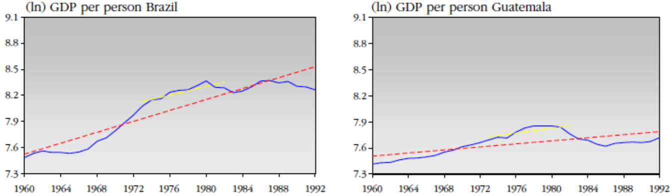


- Hills: "These 27 countries had growth rates higher than 1.5 percent in each period ... Like the United States, most of the OECD countries are hills" (but also Costa Rica and Pakistan)

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

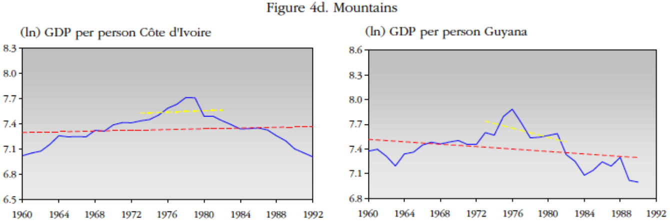
Figure 4c. Plateaus



- Plateaux: "These 16 countries had growth rates higher than 1.5 percent before their structural break, but afterward growth fell to less than 1.5 percent ... the classic case is Brazil"

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

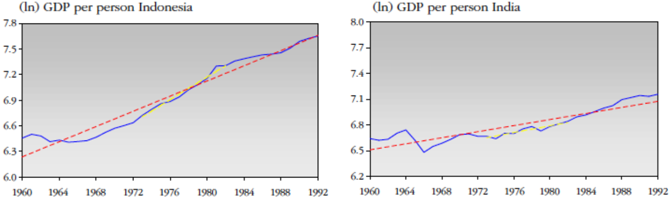


- Mountains: "These 33 countries had growth rates higher than 1.5 percent before their trend break, but negative rates afterward (figure 4d). This category includes most of the oil-exporting countries (Algeria, Gabon, Nigeria, Saudi Arabia), a number of commodity exporters that experienced positive commodity price shocks followed by negative shocks (Cote d Ivoire, Guyana, Jamaica, Zambia), and Latin American countries affected by the debt crisis (Argentina, Bolivia, Paraguay)"

Some relevant papers

Pritchett (2000). "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains." The World Bank Economic Review.

Figure 4f. Accelerators



- Accelerators: "These 7 countries did not have growth rates above 1.5 percent before their structural break, but did afterward . This class includes a number of clear successes, like Indonesia after 1966 and Mauritius after 1970, as well as less clear-cut successes, like India)"

Some relevant papers

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. population growth **(-)**
- Method of analysis: semiparametric regression

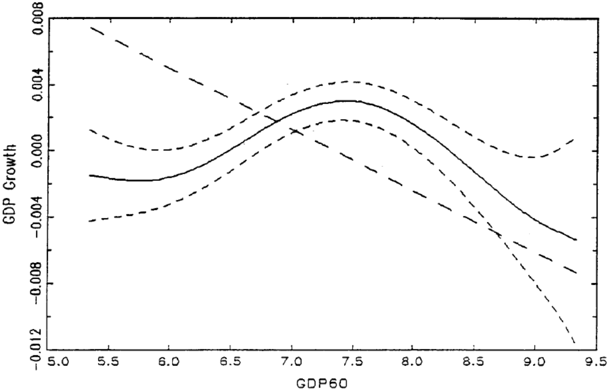
$$\gamma_i = \alpha + f_\beta(log(y_{i,0})) + \pi_n log(n_i + g + \delta) + \pi_K logs_{K,i} + f_{\pi_H}(log(s_{H,i})) + \epsilon_i$$

where $f_\beta(.)$ and $f_{\pi_H}(.)$ are arbitrary functions. The effect of $log(n_i + g + \delta)$ is estimated parametrically, the effects of $logy_{i,0}$ and $logs_{H,i}$ are estimated nonparametrically (this choice follows previous studies).

Some relevant papers

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

First finding: the effect of $log(y_{i,0})$ is negative only for incomes above \$1800



Some relevant papers

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

Second finding: 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.

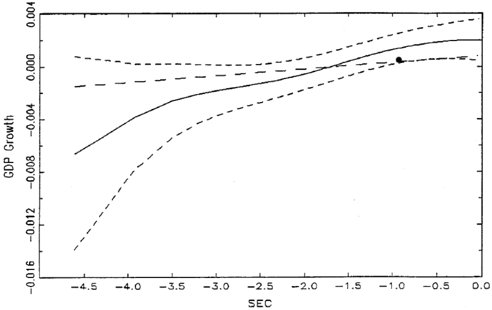


Figure 2. GDP growth versus *SEC* ($c = 1{\cdot}7$)

- This approach is 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)

- This approach is 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.

Some relevant papers

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:
  1. initial GDP
  2. initial human capital
  3. investment
  4. population growth
- 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 subgroups of countries identified by initial conditions obey distinct Solow-type regressions"

Some relevant papers

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

- 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.

Some relevant papers

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

I 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.

II 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

Some relevant papers

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

I 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.

II 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

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

Some relevant papers

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

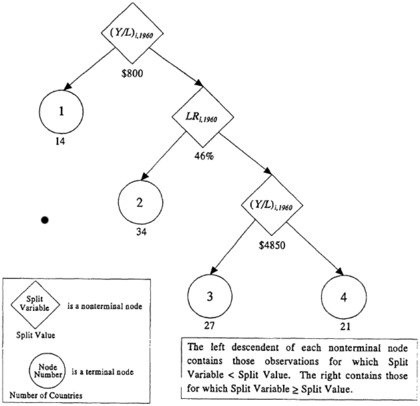


Figure 1. Regression tree

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

Terminal node number			
1	2	3	4
Burkina Faso	Algeria	Madagascar	Austria
Burundi	Angola	South Africa	Belgium
Ethiopia	Benin	Hong Kong	Denmark
Malawi	Cameroon	Israel	Finland
Mali	Central African Rep.	Japan	France
Mauritania	Chad	Korea	Federal Republic of Germany
Niger	Congo, People's Rep.	Malaysia	Italy
Rwanda	Egypt	Philippines	The Netherlands
Sierra Leone	Ghana	Singapore	Norway
Tanzania	Ivory Coast	Sri Lanka	Sweden
Togo	Kenya	Thailand	Switzerland
Uganda	Liberia	Greece	United Kingdom
Zaire	Morocco	Ireland	Canada
Burma	Mozambique	Portugal	Trinidad and Tobago
	Nigeria	Spain	United States of America
	Senegal	Costa Rica	Argentina
	Somalia	Dominican Republic	Chile
	Sudan	El Salvador	Uruguay
	Tunisia	Jamaica	Venezuela
	Zambia	Mexico	Australia
	Zimbabwe	Nicaragua	New Zealand
	Bangladesh	Panama	
	India	Brazil	
	Jordan	Columbia	
	Nepal	Ecuador	
	Pakistan	Paraguay	
	Syria	Peru	
	Turkey		
	Guatemala		
	Haiti		
	Honduras		
	Bolivia		
	Indonesia		
	Papua New Guinea		

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

- Table V. Cross-section regressions: regression tree sample breaks: dependent variable: $\ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960}$

	Terminal node number			
	1	2	3	4
Observations	14	34	27	21
	Unconstrained regressions			
Constant	3·46	−0·915	0·277	−7·26[a]
	(2·27)	(1·79)	(1·42)	(1·59)
$\ln(Y/L)_{i,1960}$	−0·791[a]	−0·086	−0·316[a]	0·069
	(0·269)	(0·131)	(0·123)	(0·139)
$\ln(I/Y)_i$	0·314[a]	0·129	1·110[a]	0·475[a]
	(0·109)	(0·159)	(0·165)	(0·119)
$\ln(n+g+\delta)_i$	−0·429	−0·390	0·059	−1·75[a]
	(0·678)	(0·489)	(0·451)	(0·270)
$\ln(SCHOOL)_i$	−0·028	0·469[a]	−0·114	0·341[a]
	(0·073)	(0·095)	(0·167)	(0·141)
$\bar{R}^2$	0·57	0·52	0·57	0·82
σ_ε	0·16	0·28	0·28	0·12
	Constrained regressions			
Θ	4·107[a]	0·539	−3·95	−11·0
	(0·552)	(1·809)	(2·67)	(7·64)
α	0·306[a]	0·186	0·758[a]	0·333[a]
	(0·083)	(0·123)	(0·095)	(0·100)
γ	−0·034	0·416[a]	−0·073	0·455[a]
	(0·083)	(0·080)	(0·114)	(0·103)
$\bar{R}^2$	0·64	0·40	0·55	0·71
σ_ε	0·19	0·32	0·30	0·18

Some relevant papers

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

- 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).

Some relevant papers

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

- 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.
- See also Desdoigts (1999) and Tan (2010)

Some relevant papers

Durlauf *et. al* (2001), "The Local Solow Growth Model", Eur. Ec Rev

- 98 countries observed in 1960-1985
- Dependent variable: growth rate
- Explanatory variables:
  1. initial GDP
  2. human capital
  3. investment
  4. population growth
- Method of analysis: estimation of a growth equation of the form:

$$\gamma_i = \alpha(y_0) + \pi_n(y_0) log(n_i + g + \delta) + \pi_K(y_0) log(s_{K,i}) + \pi_H(y_0) log(s_{H,i}) + \epsilon_i$$

Some relevant papers

Durlauf *et. al* (2001), "The Local Solow Growth Model", Eur. Ec Rev

- 98 countries observed in 1960-1985
- Dependent variable: growth rate
- Explanatory variables:
  1. initial GDP
  2. human capital
  3. investment
  4. population growth
- Method of analysis: estimation of a growth equation of the form:

$$\gamma_i = \alpha(y_0) + \pi_n(y_0)log(n_i + g + \delta) + \pi_K(y_0)log(s_{K,i}) + \pi_H(y_0)log(s_{H,i}) + \epsilon_i$$

- that is: the parameters are assumed to vary *locally*, i.e. to depend on initial income.

Some relevant papers

Durlauf *et. al* (2001), "The Local Solow Growth Model", Eur. Ec Rev

- 98 countries observed in 1960-1985
- Dependent variable: growth rate
- Explanatory variables:
  1. initial GDP
  2. human capital
  3. investment
  4. population growth
- Method of analysis: estimation of a growth equation of the form:

$$\gamma_i = \alpha(y_0) + \pi_n(y_0) log(n_i + g + \delta) + \pi_K(y_0) log(s_{K,i}) + \pi_H(y_0) log(s_{H,i}) + \epsilon_i$$

- that is: the parameters are assumed to vary *locally*, i.e. to depend on initial income.
- The method is implemented through the estimation of a local linear model in which the observations near the point where the marginal effect is estimated are weighted by a kernel function.

Some relevant papers

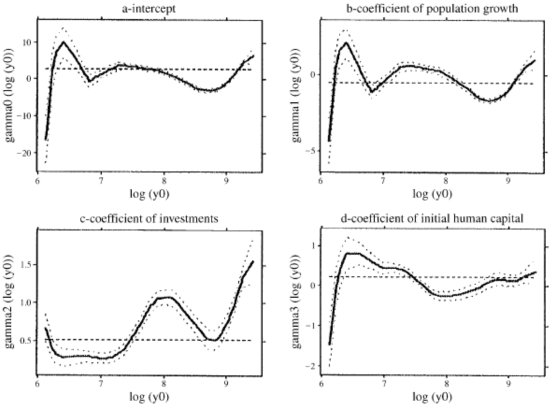
Durlauf *et. al* (2001), "The Local Solow Growth Model", Eur. Ec Rev

- 98 countries observed in 1960-1985
- Dependent variable: growth rate
- Explanatory variables:
  1. initial GDP
  2. human capital
  3. investment
  4. population growth
- Method of analysis: estimation of a growth equation of the form:

$$\gamma_i = \alpha(y_0) + \pi_n(y_0)log(n_i + g + \delta) + \pi_K(y_0)log(s_{K,i}) + \pi_H(y_0)log(s_{H,i}) + \epsilon_i$$

- that is: the parameters are assumed to vary *locally*, i.e. to depend on initial income.
- The method is implemented through the estimation of a local linear model in which the observations near the point where the marginal effect is estimated are weighted by a kernel function.
- Results: high parameter heterogeneity for poorer countries. In particular: the estimated coefficient for the intercept, population growth, and human capital stabilizes after a threshold income level; the parameter for investment is highly unstable.

Durlauf *et. al* (2001), “The Local Solow Growth Model”, EER



Reading list

- Azariadis, C. and Drazen, A. (1990), "Threshold Externalities in Economic Development", *Quarterly Journal of Economics*, 105, 501-526.
- Barro, R. J. and X. Sala-i-Martin (2004), "Economic Growth. Second Edition", MIT Press.
- Baumol, William J. (1986), "Productivity growth, convergence, and welfare: what the long-run data show", *The American Economic Review*, 1072-1085.
- Bernard, A. and S. N. Durlauf (1996), "Interpreting Tests of Convergence Hypothesis", *Journal of Econometrics*, 71, 161-173.
- 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, 365-384.
- 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.

Reading list

- Fiaschi, D. and A. M. Lavezzi (2003), "Distribution Dynamics and Nonlinear Growth", *Journal of Economic Growth*, 8, 379-401.
- Fiaschi D. e A. M. Lavezzi (2007), "Productivity polarization and sectoral dynamics in European regions", *Journal of Macroeconomics*, 29, 612-637.
- Galor, O. (1996), "Convergence? Inferences from Theoretical Models", *Economic Journal*, 106, 1056-1069 (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 Madsen, R. W. (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, 317-321.
- 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, 229-254.
- Liu, Z., and T. Stengos (1999), "Non-linearities in Cross-Country Growth Regressions: a Semiparametric Approach", *Journal of Applied Econometrics* 14, 527-538.

- 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.
- Pritchett, (2000, "Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains", *The World Bank Economic Review* 14, 221-250.
- Quah, D. T. (1993), "Empirical Cross-section Dynamics for Economic Growth", *European Economic Review* 37, 426-434.
- 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, 112-156.