World development indicators
Which country will develop more

Group 12:
Stefano Moawad
Leonardo Comandini
Diana Isaeva
Andrea Schiavon
Viktor Snesarevskii
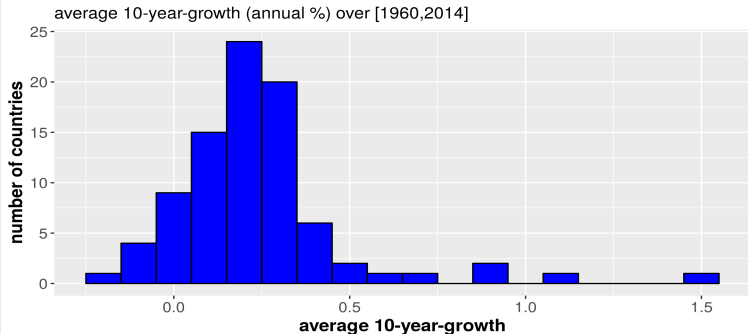
April-May 2017

kaggle

Table of Contents
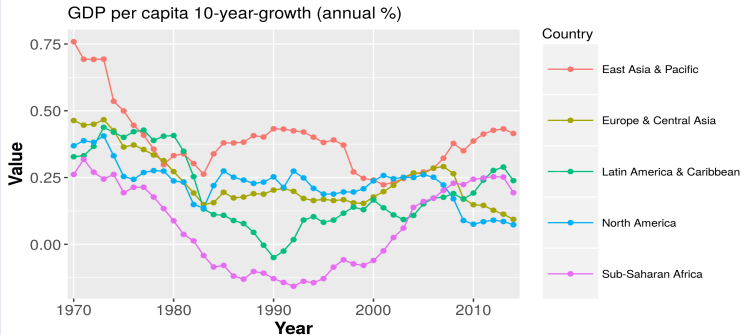
average 10-year-growth (annual %) over [1960,2014]

Definition

The 10-year-Growth is the 10-year percentage variation of the GDP per capita in local currency. More formally,
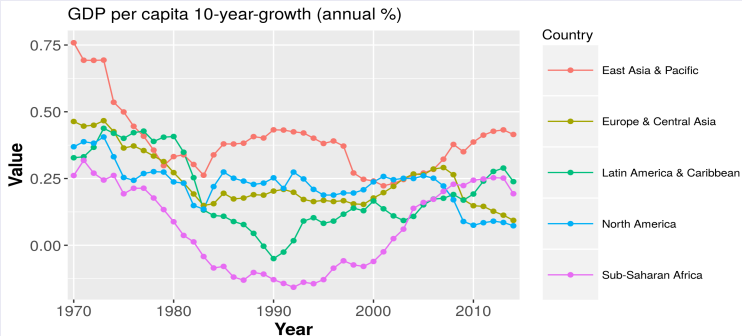
$$Growth_t := \frac{GDP_t - GDP_{t-10}}{GDP_{t-10}} \quad (1)$$

where GDP is the Gross Domestic Product per capita

10-year-growth by region

GDP per capita 10-year-growth (annual %)

10-year-growth by region

GDP per capita 10-year-growth (annual %)

Country
- East Asia & Pacific
- Europe & Central Asia
- Latin America & Caribbean
- North America
- Sub-Saharan Africa

Remark 1

significant differences between decades $\implies$ dummy for decades

GDP per capita 10-year-growth (annual %)
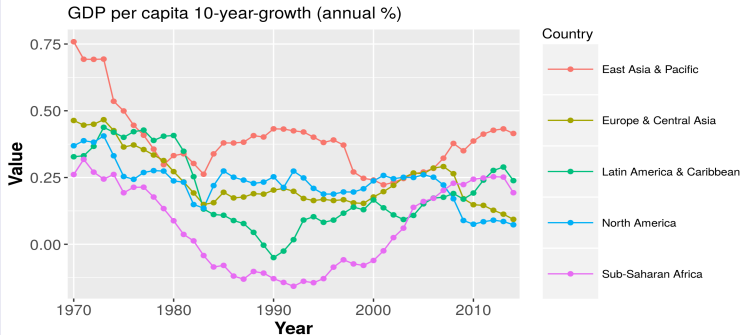
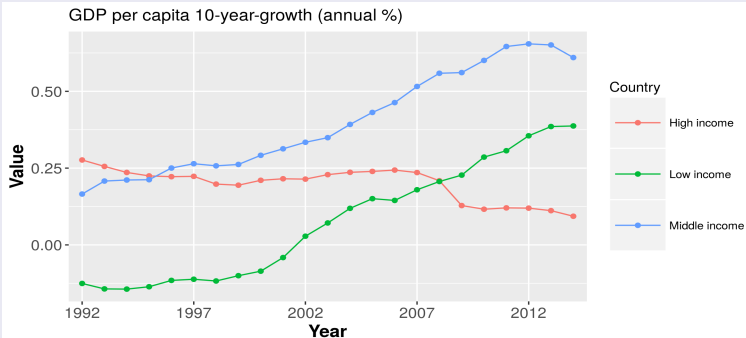**Remark 1**

significant differences between decades $\implies$ dummy for decades

Remark 2

different growth patterns for different regions $\implies$ dummy for Asia and dummy for Africa

10-year-growth by Income group

GDP per capita 10-year-growth (annual %)

Remark 3

different growth patterns for different Income groups $\implies$ dummy for High Income and dummy for Low Income

State variables

- Education $:= \frac{\text{tot enrolment primary school}}{\text{population}}$ [%]
- Health $:= \frac{1}{\text{life expectancy at birth}}$ [year]$^{-1}$
- Fertility $:=$ average number of births per woman

The Regressors: State and Environmental variables

State variables

- Education := $\frac{\text{tot enrolment primary school}}{\text{population}}$ [%]
- Health := $\frac{1}{\text{life expectancy at birth}}$ [year]$^{-1}$
- Fertility := average number of births per woman

Envirnonmental variables

- Inflation [%]
- GDP := log(GDP)
- FDI := financial capital owned by foreign investors [% of GDP]
- Openess := $\frac{\text{Inport+Export}}{\text{GDP}}$
- Consumption := households consumption expenditure [% of GDP]
- Investment := government expenditures for goods and services [% of GDP]

Let $X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{3n}]^T$ be the design matrix

Dummies for decades

$$D_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1983, 1993) \\ 0 & \text{otherwise} \end{cases} \qquad D_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1993, 2003) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Let $X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{3n}]^T$ be the design matrix

Dummies for decades

$$D_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1983, 1993) \\ 0 & \text{otherwise} \end{cases} \qquad D_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1993, 2003) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Dummies for Income

$$I_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{High Income} \\ 0 & \text{otherwise} \end{cases} \qquad I_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{Low Income} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

The Dummies: decade, income and region

Let $X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{3n}]^T$ be the design matrix

Dummies for decades

$$D_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1983, 1993) \\ 0 & \text{otherwise} \end{cases} \qquad D_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in [1993, 2003) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Dummies for Income

$$I_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{High Income} \\ 0 & \text{otherwise} \end{cases} \qquad I_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{Low Income} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Dummies for Region

$$R_1 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{East Asia} \\ 0 & \text{otherwise} \end{cases} \qquad R_2 = \begin{cases} 1 & \text{if } \underline{x}_i \in \text{Sub-Saharan Africa} \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

Let $\epsilon \sim N(0, \sigma^2)$

complete model

$$
\begin{aligned}
Growth = {} & \beta_0 + \beta_1 \text{fertility} + \beta_2 \text{FDI} + \beta_3 \text{GDP} + \beta_4 \text{education} + \beta_5 \text{consumption} + \\
& \beta_6 \text{inflation} + \beta_7 \text{health} + \beta_8 \text{investment} + \beta_9 \text{openess} + \\
& \beta_{10} D_1 + \beta_{11} D_2 + \beta_{12} R_1 + \beta_{13} R_2 + \beta_{14} I_1 + \beta_{15} I_2 + \\
& D_1(\beta_{16} \text{GDP} + \beta_{17} \text{fertility} + \beta_{18} \text{investment}) + \\
& D_2(\beta_{19} \text{GDP} + \beta_{20} \text{fertility} + \beta_{21} \text{investment}) + \\
& I_1(\beta_{22} \text{GDP} + \beta_{23} \text{fertility} + \beta_{24} \text{investment}) + \\
& I_2(\beta_{25} \text{GDP} + \beta_{26} \text{fertility} + \beta_{27} \text{investment}) + \\
& R_1(\beta_{28} \text{GDP} + \beta_{29} \text{fertility} + \beta_{30} \text{investment}) + \\
& R_2(\beta_{31} \text{GDP} + \beta_{32} \text{fertility} + \beta_{33} \text{investment}) + \epsilon
\end{aligned}
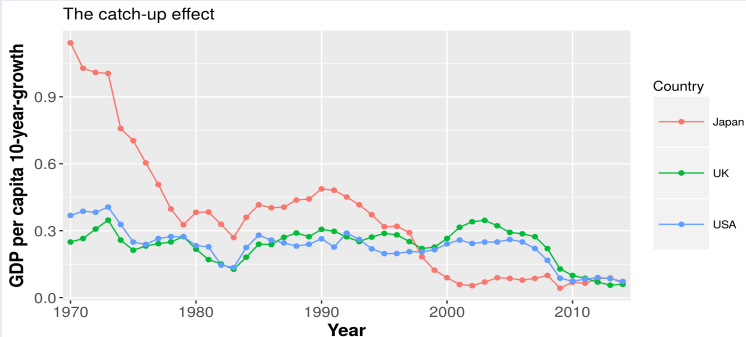$$

Let $\epsilon \sim N(0, \sigma^2)$

reduced model

$$
\begin{aligned}
Growth = {} & \beta_0 + \beta_1 \text{fertility} + \beta_2 \text{GDP} + \beta_3 \text{consumption} + \beta_4 \text{investment} + \\
& \beta_5 \text{education} + \beta_6 \text{FDI} \\
& \beta_7 D_1 + \beta_8 D_2 + \beta_9 R_1 + \beta_{10} R_2 + \beta_{11} I_1 + \beta_{12} I_2 + \\
& D_1(\beta_{13} \text{GDP} + \beta_{14} \text{investment}) + \\
& D_2(\beta_{15} \text{GDP}) + \\
& I_1(\beta_{16} \text{GDP} + \beta_{17} \text{fertility} + \beta_{18} \text{investment}) \\
& I_2(\beta_{19} \text{investment}) + \\
& R_1(\beta_{20} \text{GDP} + \beta_{21} \text{fertility}) + \\
& R_2(\beta_{22} \text{fertility} + \beta_{23} \text{investment}) + \epsilon
\end{aligned}
$$

lm output

	Model 1
(Intercept)	0.9531 (0.3791)[*]
fertility	−0.0849 (0.0244)[***]
FDI	−0.0085 (0.0063)
GDP	−0.0903 (0.0305)[**]
education	−0.0025 (0.0010)[*]
consumption	0.0047 (0.0010)[***]
health	−21.0428 (11.8102)
R1	3.8459 (0.3718)[***]
R2	0.8626 (0.1585)[***]
I1	1.0546 (0.4503)[*]
I2	−0.4912 (0.1445)[**]
investment	0.0407 (0.0063)[***]
D1	−0.3408 (0.1521)[*]
D2	−0.4913 (0.1369)[***]
GDP:D1	0.0841 (0.0182)[***]
investment:D1	−0.0189 (0.0048)[***]
GDP:D2	0.0640 (0.0163)[***]
GDP:I1	−0.0783 (0.0474)
fertility:I1	0.0804 (0.0297)[**]
investment:I1	−0.0354 (0.0073)[***]
investment:I2	0.0327 (0.0088)[***]
GDP:R1	−0.3070 (0.0348)[***]
fertility:R1	−0.3880 (0.0362)[***]
fertility:R2	−0.0527 (0.0274)
investment:R2	−0.0425 (0.0073)[***]

R^2	0.8705
Adj. R^2	0.8364
Num. obs.	116
RMSE	0.1134

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Legend:
- D1 = [1983,1993]   D2 = [1993,2003]
- R1 = Asia   R2 = Africa
- I1 = high income
- I2 = low income

The catch-up effect

Japan, UK, USA — GDP per capita 10-year-growth vs Year

Definition

the following principle is called **conditional convergence**: the lower the initial GDP the higher the growth over the next decade

Let $\epsilon \sim N(0, \sigma^2)$

prediction model: timeless

$Growth = \beta_0 + \beta_1 \text{fertility} + \beta_2 \text{GDP} + \beta_3 \text{consumption} + \beta_4 \text{investment} +$
$\qquad \beta_5 R_1 + \beta_6 R_2 + \beta_7 I_1 + \beta_8 I_2 +$
$\qquad I_1(\beta_9 \text{GDP} + \beta_{10} \text{fertility} + \beta_{11} \text{investment})$
$\qquad I_2(\beta_{12} \text{investment} + \beta_{13} \text{fertility}) +$
$\qquad R_1(\beta_{14} \text{GDP} + \beta_{15} \text{fertility}) +$
$\qquad R_2(\beta_{16} \text{fertility} + \beta_{17} \text{investment}) + \epsilon$

predictor evaluation

fitting sample = [1983,2003]   test sample = [2003,2013]

Definition

- $F_t =$ prediction for the growth in t with our model
- $Y_t =$ realization of growth in t
- $e_t =$ prediction error
- $ME = \sum_{t=0}^{T} e_t =$ mean error over $[t, T]$
- $MAD = \sum_{t=0}^{T} \|e_t\| =$ mean absolute deviation over $[t, T]$
- $RMSE = \sqrt{\sum_{t=0}^{T} e_t^2} =$ root mean square error

predictor evaluation

fitting sample = [1983,2003] test sample = [2003,2013]

Definition

- F_t = prediction for the growth in t with our model
- Y_t = realization of growth in t
- e_t = prediction error
- $ME = \sum_{t=0}^{T} e_t$ = mean error over $[t, T]$
- $MAD = \sum_{t=0}^{T} \|e_t\|$ = mean absolute deviation over $[t, T]$
- $RMSE = \sqrt{\sum_{t=0}^{T} e_t^2}$ = root mean square error

predictor performances

	same countries	new countries
ME	-0.014	0.022
MAD	0.099	0.182
RMSE	0.1361	0.2823

- slightly overestimating
- inaccurate out-of-sample

prediction 2020 GDP per capita 10-year-growth
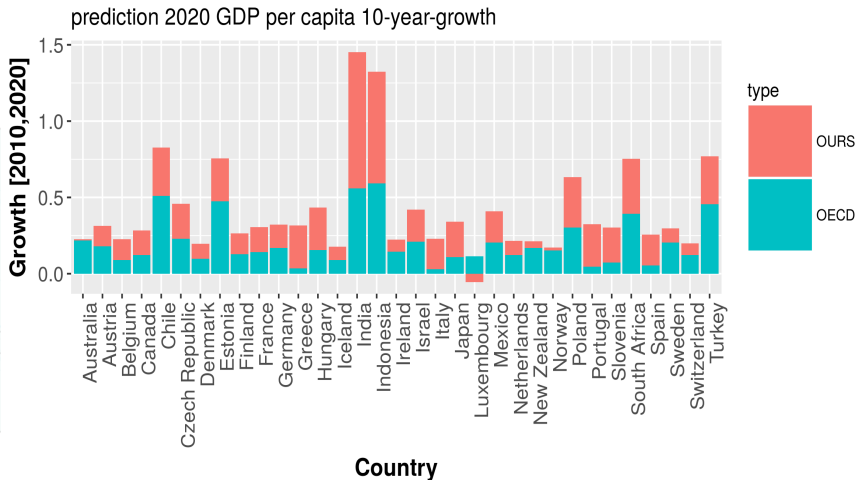
OECD = The Organisation for Economic Co-operation and Development is an intergovernmental economic organisation with 35 member countries, founded in 1960 to stimulate economic progress and world trade