Seminar to Advanced Macroeconomics

Does the income of poor countries converge to the rich ones?

Do countries with lower initial income grow faster?

Outline

- Convergence of income: unconditional and conditional one
- Is the convergence hypothesis supported by the data?
- Note: Omitted variable bias
- Extension: panel data

Convergence

- Rationale for convergence of income:
  - Diminishing returns in neoclassical models implie that countries with lower initial income will grow faster.
  - Therefore, poor and rich countries should converge in terms of income levels per capita.
- Is this convergence-hypothesis supported with the data?

Convergence

- Regress output growth over some period on a constant and initial income (Unconditional convergence)

$$\log\left(\frac{Y}{L}\right)_{i,t=2} - \log\left(\frac{Y}{L}\right)_{i,t=1} = a + b\,\log\left(\frac{Y}{L}\right)_{i,t=1} + \epsilon_i$$

- Problems => lack of data and also their reliability namely at t=1. (availability of data for poor countries before 1960)

Convergence - Results

Model 1: OLS, using observations 1-75
Dependent variable: loggrowth

	coefficient	std. error	t-ratio	p-value
const	0.587524	0.432926	1.357	0.1789
l_gdp60	-0.00423507	0.0548352	-0.07723	0.9386

- Very poor results, negative convergence rate identified.
- Sample sensitive: years and selection of countries does matter, that's why in Romer's textbook slightly different results
- Why: poor growth in 80's in most developing countries etc.

Let's put it in a different way...

- Barro (1989):
  - In neoclassical growth models with diminishing returns such as Solow (1956) and Koopmans (1965), a country's per capita growth rate tends to be invesely related to its starting level of income per person. Therefore, in the absence of shocks, poor and rich countries would tend to converge in terms of levels of per capita income. However this convergence hypothesis seems to be inconsistent with the cross country evidence, which indicates that per capita growth rates are uncorrelated with the starting level of per capita output.
- Mankiw-Romer-Weil (1992):
  - Countries might have different steady states but if we control for the determinants of steady state (namely the saving rates), *conditional* convergence occurs.

Convergence and Solow model

- Conditional convergence:
  - Regress output growth on initial income and saving rates.

$$\log y(t) - \log y(0) = \beta_1 \log s_k + \beta_2 \log(n + g + \delta) + \beta_3 \log y_0$$

$$\log y(t) - \log y(0) = \beta_1 \log s_k + \beta_2 \log(n + g + \delta) + \beta_3 \log s_h + \beta_4 \log y_0$$

Conditional Convergence - Results

Model 2: OLS, using observations 1-75
Dependent variable: loggrowth

	coefficient	std. error	t-ratio	p-value	
const	3.709000	0.908784	4.081	0.0001	***
l_gdp60	-0.365993	0.0674345	-5.427	7.76e-07	***
s	0.537563	0.102293	5.255	1.52e-06	***
ngd	-0.544986	0.288428	-1.890	0.0630	*
ls	0.270455	0.0803728	3.365	0.0012	***

- Much better results.
- Growth is negatively correlated with initial income if we controll for other variables.

Script

```
open /.../mrw.gdt
logs gdp85
genr s = log(inv/100)
genr ngd = log(popgrow/100+0.05)
smpl nonoil –dummy
smpl intermed --dummy
# model 1
ols l_gdp85 const s ngd
genr ls=log(school/100)
# model 2
ols l_gdp85 const s ngd ls
genr LogGrowth=l_gdp85-l_gdp60
# model 3
ols LogGrowth const l_gdp60
# model 4
ols LogGrowth const l_gdp60 s ngd
# model 5
ols LogGrowth const l_gdp60 s ngd ls
```

Note: Omitted variable bias

- Econometric point of view: convergence appeared to be present in the data after inclusion of several other variables.
- So called *omitted variable bias*: if any variable is not included in the model, it's effect is hidden in residuals. But if this variable is also correlated with explanatory variables (like here with initial GDP), then residuals are correlated with regressors and an assumption of OLS of cov(u,X) = 0 is violated. Estimates are then biased.

Note: Omitted variable bias

- Formal deviation (see wikipedia for details) http://en.wikipedia.org/wiki/Omitted_variable_bias

$$y_i = x_i\beta + z_i\delta + u_i, \qquad i = 1, \ldots, n$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta + Z\delta + U) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'Z\delta + (X'X)^{-1}X'U \\ &= \beta + (X'X)^{-1}X'Z\delta + (X'X)^{-1}X'U.\end{aligned}$$

$$\begin{aligned}E[\hat{\beta}] &= \beta + (X'X)^{-1}X'Z\delta \\ &= \beta + \text{bias}.\end{aligned}$$

Note: Omitted variable bias Practitioner's point of view

- How to detect omitted variable bias?
  - No single test, but sensitivity analysis: include several other variables that are supposed to might matter (either in one regression with all variables or in a set of regressions with only one additional variable included) – if sign and size of the coefficient at variable-of-interest (here the y_0) consistent and in line with economic intuition, then the results are most probably correct.
- How to prevent it?
  - Use economic theory to establish a set of alternative explanatory variables (Well, this is the practical usage of theories, that we are trying to show you... Consider the link Solow model – Growth regression).
  - Follow existing research papers and check for the results and variables included there (the authors usually tried a lot of stuff so why not using existing knowledge).
  - Do sensitivity analysis mentioned above.

Panel Data Analysis

- Introduction into the panel data analysis before we use them on the next seminar
- Types of data in econometrics:
  - Cross sectional (data across countries in one year/quarter)
  - Time series (time series of evolution of the GDP in one country)
  - Panel data: mix of both dimensions (several variables observed in a number of countries and in several points of times)
- Example: Data set from the Penn World Tables, see gretl sample files => Gretl => penngrow
- Contains variables for testing the Solow model not only for one or two years, but for six.

Typical panel

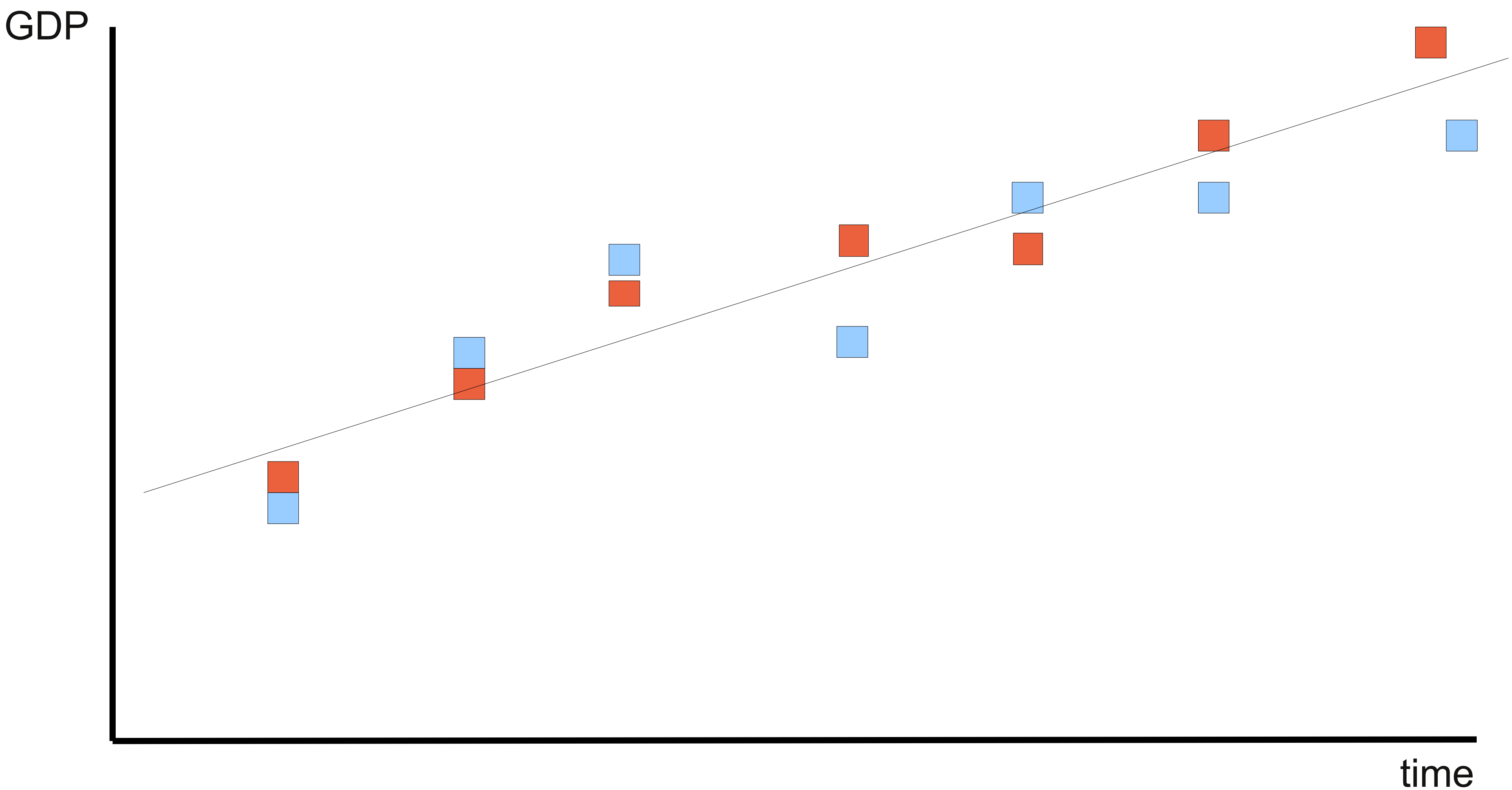
	country	year	Y	X
1:1	53	1960	7.987185	
1:2	53	1965	8.403801	1.62254
1:3	53	1970	8.896177	1.76904
1:4	53	1975	9.033126	1.78886
1:5	53	1980	9.217117	1.78772
1:6	53	1985	9.373394	1.73075
2:1	70	1960	8.544225	
2:2	70	1965	8.720297	1.46819
2:3	70	1970	8.923191	1.57635
2:4	70	1975	9.097508	1.65637
2:5	70	1980	9.259035	1.69168
2:6	70	1985	9.317489	1.54990
3:1	71	1960	8.606851	
3:2	71	1965	8.814776	1.49272
3:3	71	1970	9.026418	1.60234
3:4	71	1975	9.171703	1.56582
3:5	71	1980	9.314340	1.56990
3:6	71	1985	9.331230	1.31784
4:1	73	1960	8.814330	
4:2	73	1965	9.037890	1.55846
4:3	73	1970	9.175956	1.63399
4:4	73	1975	9.233080	1.66633
4:5	73	1980	9.335474	1.59107
4:6	73	1985	9.470317	1.40232

...

Working with panel data

- Three ways:
  - Study each time series (or several of them) using methods of time series analysis
  - The heterogeneity of observations might be omitted and all observations can be used as if no structural differences across countries (individuals etc.) exist – so called pooled model, basically OLS regression.
  - Some method that assume comovement and allows for inividual structural differences persistent over time.
- Several methods for analysis of panel data, some of them introduced in the Econometrics A class later in the winter semester.
- Here we use so called *Fixed effects model*: individual intercepts.

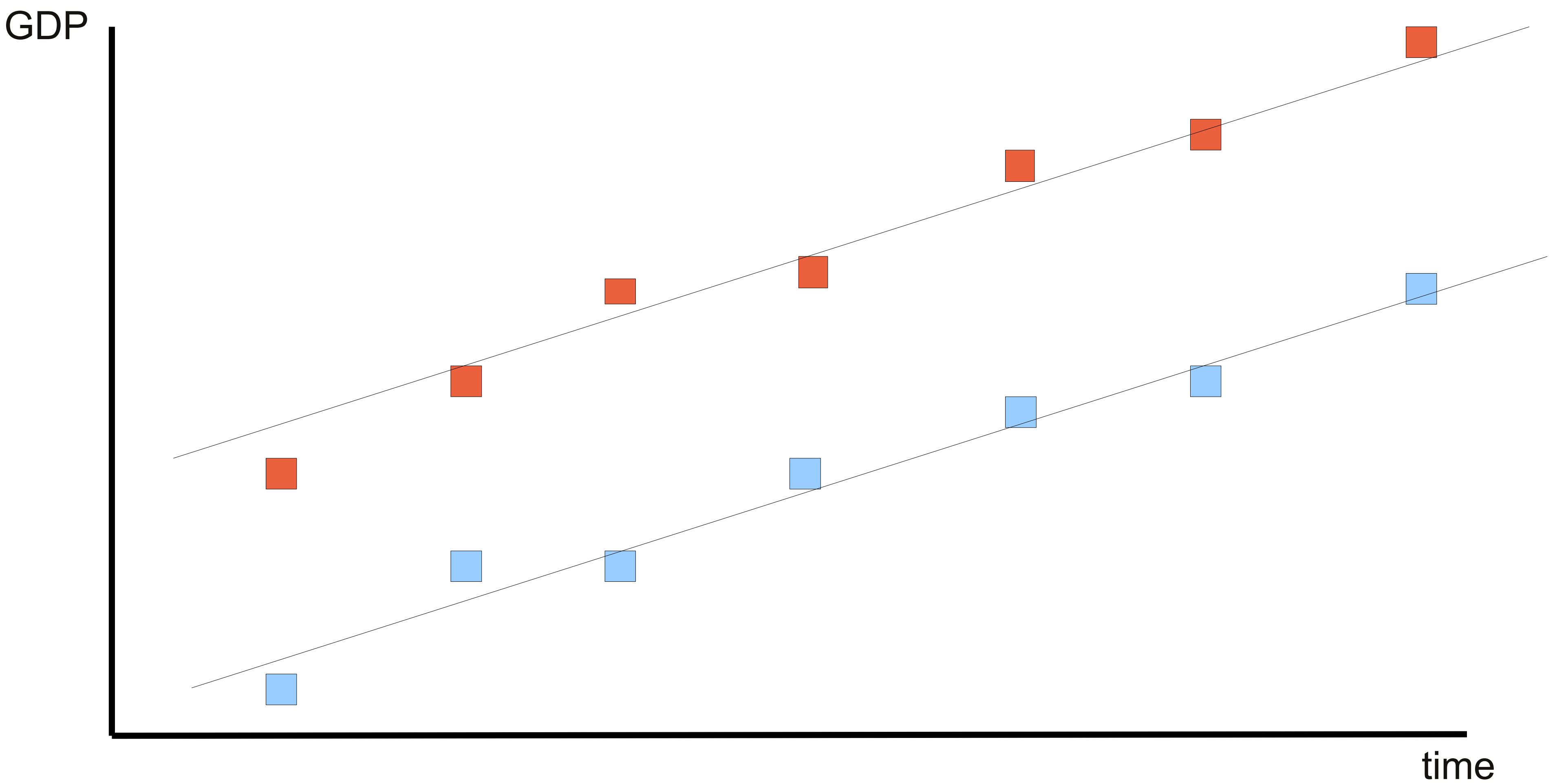
Illustration – OLS good enough



Countries RED and BLUE, model $GDP_{i,t} = c_1 + c_2 * time_{i,t} + u_{i,t}$

Here OLS seems to be adequate, countries are similar (pooled model)

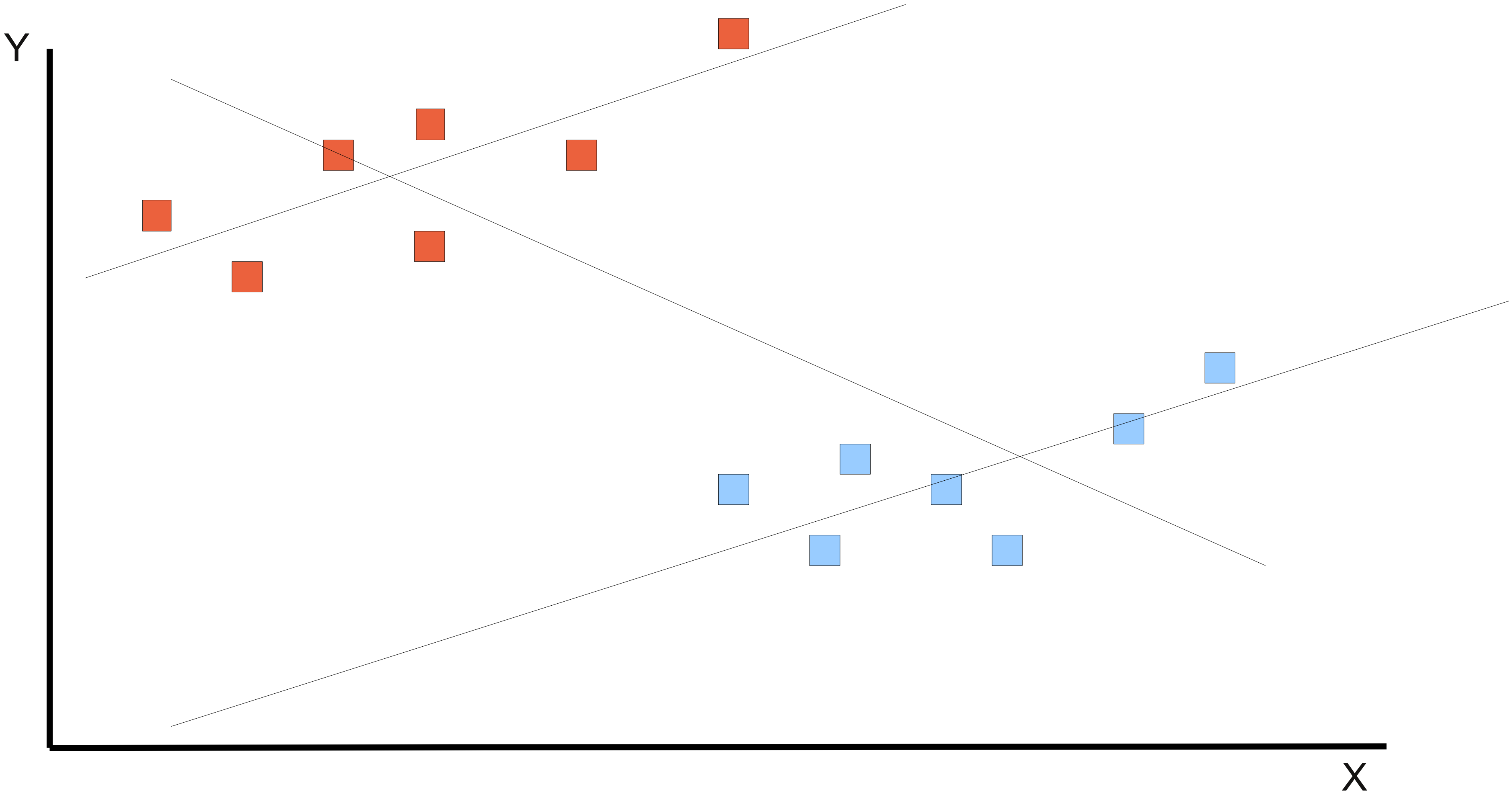
Illustration – Panel data necessary 1



Countries RED and BLUE, model $GDP_{i,t} = c_1 + c_2 * time_{i,t} + u_{i,t}$

Here OLS not adequate anymore, because two regression lines are distant

Illustration – Panel data necessary 2



Artifitial illustration of two nonhomogenous groups, where joint regression line predicts opposite relation between factors X and Y.

Working with panel data

Fixed effects model:

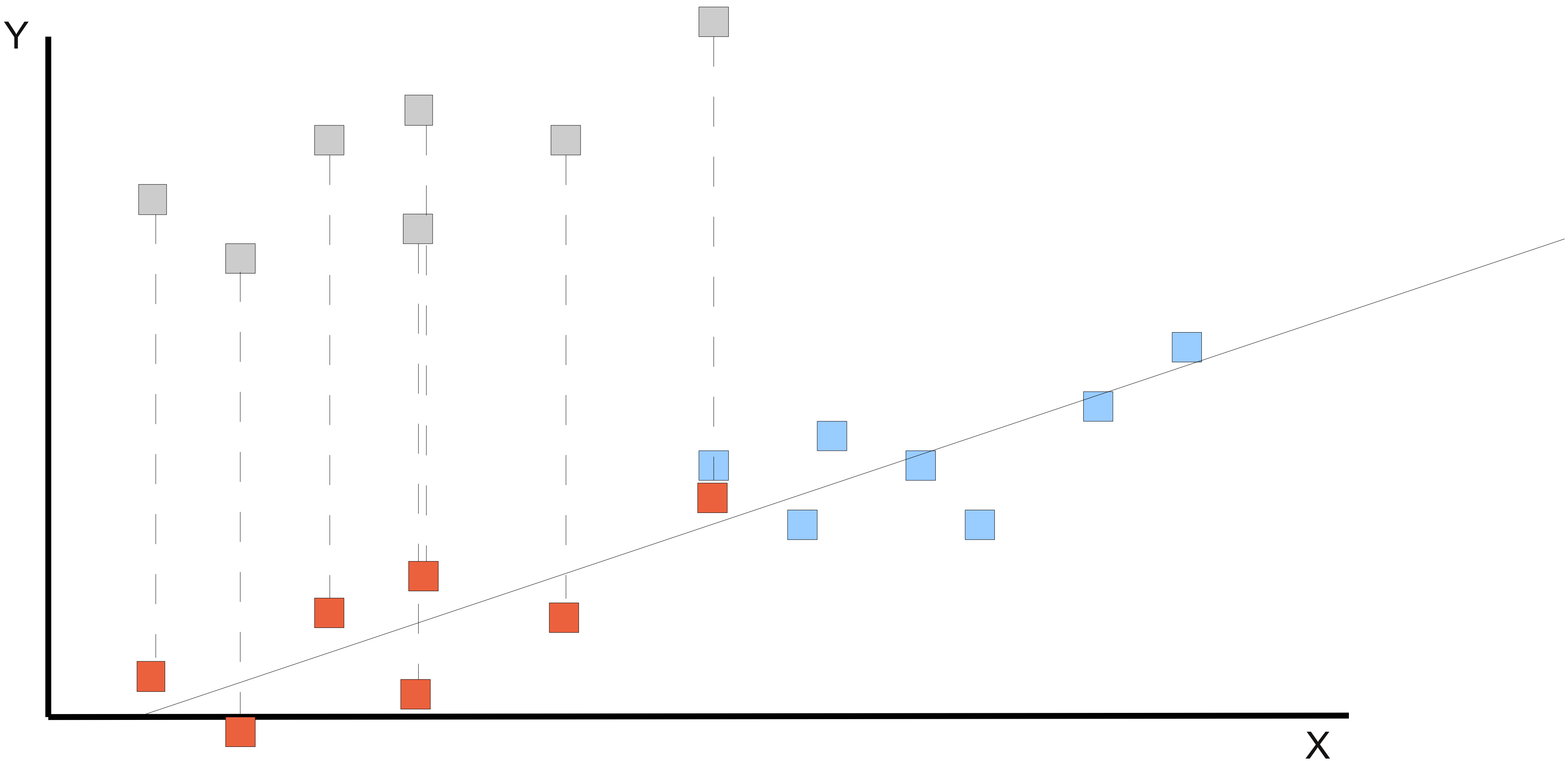
$$Y_{i,t}=\alpha+\beta_{i,t}X_{i,t}+\epsilon_{i,t}$$

translates into

$$Y_{i,t}=\alpha_i+\beta_{i,t}X_{i,t}+\epsilon_{i,t}$$

which is an equivalent to a model with dummy variable assigned to all but the first group of observations.

Illustration – Panel data necessary 2



Artifitial illustration of two nonhomogenous groups, where joint regression line predicts opposite relation between factors X and Y.

Working with panel data

Fixed effects model:

$$Y_{i,t}=\alpha+\beta_{i,t}X_{i,t}+\varepsilon_{i,t}$$

translates into

$$Y_{i,t}=\alpha_i+\beta_{i,t}X_{i,t}+\varepsilon_{i,t}$$

which is an equivalent to a model with dummy variable assigned to all but the first group of observations. However, there are number of computational issues like a large number of dummy variables with new unknown coefficients, hence a small computational trick needs to be applied: time-demeaning.

We substract a group mean from each of variables...

Define $\tilde{Y}_{i,t}=Y_{i,t}-\bar{Y}_{i,t}$ and $\tilde{X}_{i,t}=X_{i,t}-\bar{X}_{i,t}$,

where $\bar{Y}_{i,t}$ is an average value of Y in group i.

Then we etimate a model without a constant.

$$\tilde{Y}_{i,t}=\beta_{i,t}\tilde{X}_{i,t}+\varepsilon_{i,t}$$

Example

```
Model 4: Fixed-effects, using 470 observations
Included 94 cross-sectional units
Time-series length = 5
Dependent variable: Y

            coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       7.55393      0.0239101    315.9     0.0000   ***
  X           0.253360     0.0355880      7.119   5.57e-12 ***

Test for differing group intercepts -
  Null hypothesis: The groups have a common intercept
  Test statistic: F(93, 375) = 44.3278
  with p-value = P(F(93, 375) > 44.3278) = 5.93194e-156
```

Script

```
open => penngrow.gdt
# model 4
panel Y const X
```

Key points

- Conditional vs. unconditional convergence
- Omitted variable bias
- Panel data
- Fixed effects model

Note to problem sets

- Your solution should contain:
  - Few words about problem which you solve
  - Specification of model (=equation)
  - Method (usually OLS) and software used
  - Estimation results
  - Some diagnostics (test of homoscedasticity, normality and non-autocorrelated residuals, comments to significance of variables).
  - Interpretation of results.