Causes of Cross-country Income Gaps

Prof. Lutz Hendricks

Econ520

January 19, 2017

Objectives

- We start looking into the question: Why are some countries rich and others poor?
- We think about **methods** that could be used to answer such questions.

Why Are Some Countries Rich and Others Poor?

Fact: Rich countries are 25 times richer than poor countries.

What do poor countries lack?

Some candidates...

Methods

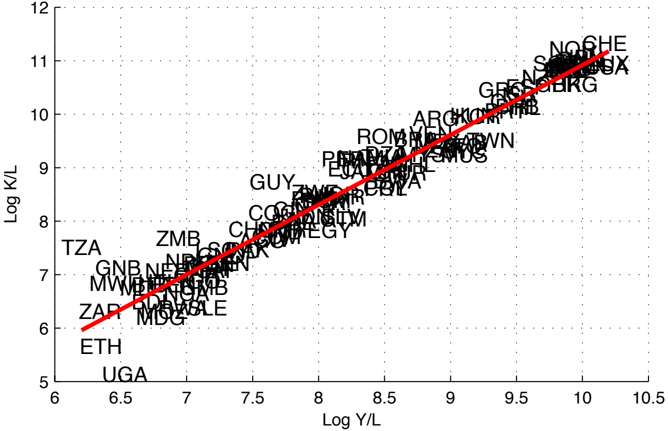
What methods could be used to answer questions such as:

How important is capital for cross-country income differences?

- Regression analysis (we will look at this one next)
- Others?

Regression Analysis

GDP and Capital Stock: 1990 data



Regression Analysis

- The figure above suggests that output and capital are related by a linear (straight line) relationship.
- We could postulate the (statistical) model:

$$\log(Y_i/L_i) = \alpha + \beta \log(K_i/L_i) + \varepsilon_i \tag{1}$$

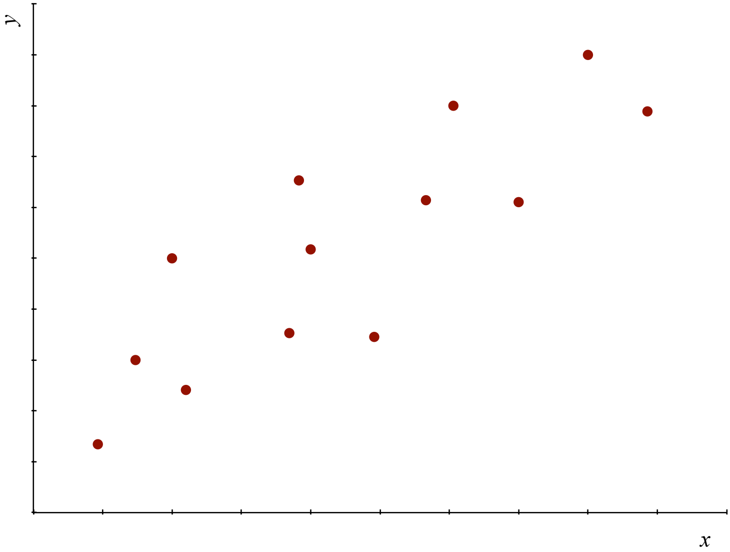
- i indexes the country
- The model "explains" the $\alpha + \beta \log(K_i/L_i)$ part of the variation in $\log(Y_i/L_i)$.
- ε_i is the unexplained **residual** – everything we have not modeled.

Ordinary Least Squares (OLS)

- OLS is a method for fitting a line through the data.
- OLS finds the coefficients (α, β) that minimize the sum of squared residuals.
- Formally, OLS solves:

$$\min_{\alpha,\beta} \sum_i (\varepsilon_i)^2 = \min \sum_i \left(\log\left(Y_i/L_i\right) - \alpha - \beta \log\left(K_i/L_i\right)\right)^2 \quad (2)$$

OLS Illustration



Multiple regression

- Typically one would add "covariates" to a regression.
- The idea is to "hold constant" other things.
- The model

$$\log(Y_i/L_i) = \alpha + \beta \log(K_i/L_i) + \sum_k \gamma_k X_{ik} + \varepsilon_i \qquad (3)$$

- X_{ik} is the value of regressor k for country i
- More compact: $\log(Y/L) = X\beta + \varepsilon$
  where Y/L and ε are $N \times 1$ vectors and X is an $N \times K$ matrix
- Examples:

Example

Table 2
Education as Determinant of Growth of Income per Capita, 1960−2000

	Dependent variable: average annual growth rate in GDP per capita, 1960–2000			
	(1)	(2)	(3)[a]	(4)
GDP per capita 1960	−0.379	−0.302	−0.277	−0.351
	(4.24)	(5.54)	(4.43)	(6.01)
Years of schooling 1960	0.369	0.026	0.052	0.004
	(3.23)	(0.34)	(0.64)	(0.05)
Test score (mean)		1.980	1.548	1.265
		(9.12)	(4.96)	(4.06)
Openness				0.508
				(1.39)
Protection against expropriation				0.388
				(2.29)
Constant	2.785	−4.737	−3.701	−4.695
	(7.41)	(5.54)	(3.32)	(5.09)
N	50	50	50	47
R^2 (adj.)	0.252	0.728	0.741	0.784

Notes: *t*-statistics in parentheses.

Source: Hanushek and Woessman (2008)

Reading a Regression Table

A made-up example:

$$Y = \alpha + \beta X + \varepsilon \tag{4}$$

$$= \underset{(0.012)}{0.123} + \underset{(0.45)}{2.34} X + \varepsilon \tag{5}$$

Point estimate $\hat{\beta} = 2.34$: the estimated "effect" of the regressor on the dependent variable

Standard error of $\hat{\beta}$ in parentheses (0.45)

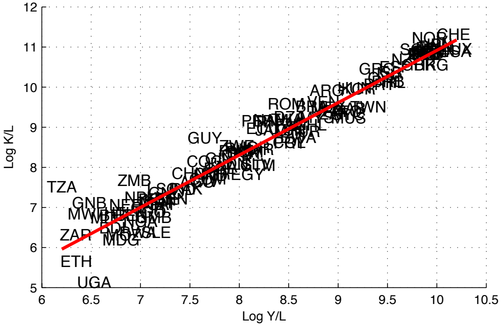
- what does that mean?

R^2: measure of "fit"

- $R^2 = 1 -$ [residual sum of squares] / [total sum of squares]
- fraction of (squared) variation in Y that is "explained" by the regression

Application to capital and output

The OLS estimate of β is about 0.5.



Just eyeballing the figure shows: variation in capital "explains" almost the entire variation in Y/L.

Are we done?

Regression Analysis: Interpretation Issues

Interpreting Regression Results

Suppose we regress

$$\ln(Y/L) = \alpha + \beta \ln(K/L) + \varepsilon \tag{6}$$

and find $\beta = 0.5$

What does $\beta = 0.5$ mean in words?

What do we learn about the question:

> *By how much would Y/L rise, if we gave a country 10% more K/L?*

Interpreting Regression Results

Key point

The OLS regression has nothing to say about this question.

Is there an easy way to prove this?

Regressions Do Not Answer Causality Questions

Proof: I can run the regression in reverse:

$$\log(K_i/L_i) = \hat{\alpha} + \hat{\beta}\log(Y_i/L_i) + \hat{\varepsilon}_i \tag{7}$$

I will get something close to $\hat{\beta} = 1/\beta = 2$.

Either regression is equally valid.

This means the regression says nothing about whether K causes Y or the other way around (or neither).

Omitted Variables

Any relevant variable omitted from the regression leads to biased results.

Example

Output depends on capital and schooling

$$\log(Y_i/L_i) = \alpha + \beta_k \log(K_i/L_i) + \beta_s s_i + \varepsilon_i \tag{8}$$

We regress output on capital only (schooling is omitted)

Result: the coefficient on capital is too large: $\hat{\beta}_k > \beta_k$

Why? Under what conditions?

Interpretation issues

Fact

OLS does nothing more than describe the data.

OLS answers the question:

> *If two observations differ by a given x, by how much do their y's differ **on average**?*

This has nothing to do with causality.

We learn nothing about the question:

> *If Greece increased its K/L by 10%, by how much would Y/L increase?*

Interpretation Issues

Fact

No statistical method can answer cause-effect questions.

A partial exception: Instrumental Variables (IV)

Instrumental Variables

The idea: find variation in the regressor that is caused by other regressors.

Suppose

$$\log(Y_i/L_i) = \alpha + \beta_k \log(K_i/L_i) + X\gamma + \varepsilon \tag{9}$$

where we don't know the covariates X.

But we also have

$$\log(K_i/L_i) = \delta + \beta_z z_i + \varepsilon_i \tag{10}$$

Assume that z has no direct effect on output (it is not part of X)

Instrumental Variables

Then the following works:

1. Regress $\log(K_i/L_i)$ on $z_i \rightarrow \hat{\beta}_z$.
2. Predict $\log\left(\widehat{K_i/L_i}\right) = \hat{\delta} + \hat{\beta}_z z_i$.
3. Regress

$$\log(Y_i/L_i) = \alpha + \beta_k \log\left(\widehat{K_i/L_i}\right) + \varepsilon_i \tag{11}$$

The resulting $\hat{\beta}_k$ is an unbiased estimator of β_k.

Example Instruments

For capital:

- natural disasters
- IMF loans

For institutions:

- institutions put in place in colonial times

For inflows of migrants:

- Mariel boat lift (Cuba)
- Refugee crisis in Syria

IV Intuition

Why does IV work?

- It makes the regressor orthogonal to all omitted regressors.
- Similar to a natural experiment

Key: one must be able to argue that the instrument has no direct effect on the regressand (output).

It is never possible to prove this.

Validity of an instrument is a subjective judgement.

How Can We Answer Cause/Effect Questions?

Possible methods:

1. controlled experiments
   almost never possible in economics
2. natural experiments (see below)
   these are rare
3. case studies
   subject to interpretation issues
4. instrumental variables
5. quantitative models

Natural Experiments

This is as close as we can get to experimental evidence in social sciences.

The idea:

By a fluke of nature, something varies "at random" across countries

Examples?

Summary

Statistical methods can **describe** data (useful).

- e.g.: capital and output are highly correlated across countries

They cannot answer **cause-effect** questions

- e.g.: by how much would output rise, if we gave a country more capital?

How can we answer cause-effect questions?

- natural experiments (rare)
- quantitative models

Reading

A good reference for econometrics (practical issues and interpretation) is Kennedy (2008).

The blog entry Against Multiple Regression and the interview it points to highlight the limitations of regression analysis.

The intuition underlying Instrumental Variables is explained here.

References I

Hanushek, E. A. and Woessman, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, **46** (3), 607–668.

Kennedy, P. (2008). *A Guide to Econometrics. 6th edition*. Wiley-Blackwell, 6th edn.