Causes of Cross-country Income Gaps

Prof. Lutz Hendricks

Econ520

October 21, 2021

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

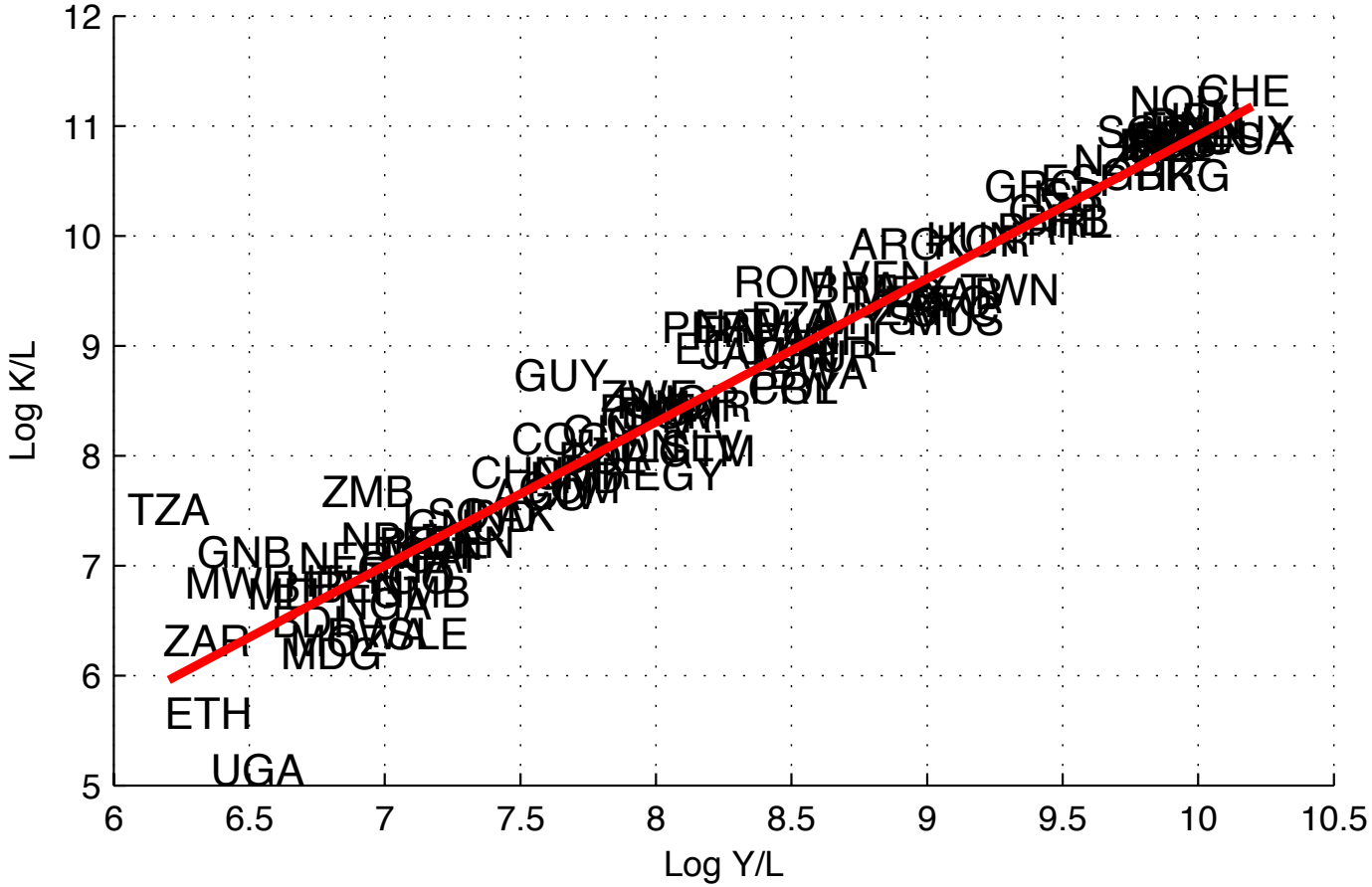
What methods could be used to answer questions such as:

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



Do differences in capital stocks cause output gaps?

What could go wrong with this interpretation?

Omitted Variables

wrong

graph +++

Source: Jones and Vollrath (2013)

Rich countries also have high human capital.

Regression Analysis

We could postulate the (statistical) model:

$$\log(Y_i/L_i) = \alpha + \underbrace{\beta \log(K_i/L_i) + \gamma H_i}_{\text{"explained"}} + \underbrace{\varepsilon_i}_{\text{residual}} \tag{1}$$

- i indexes the country
- H_i is a measure of human capital

The model "explains" part of the variation in $\log(Y_i/L_i)$.

ε_i is the unexplained **residual** – everything we have not modeled.

Next task: estimate β and γ.

Ordinary Least Squares (OLS)

OLS is a method for fitting a line through the data.

OLS finds the coefficients (α, β, γ) that **minimize the sum of squared residuals**.

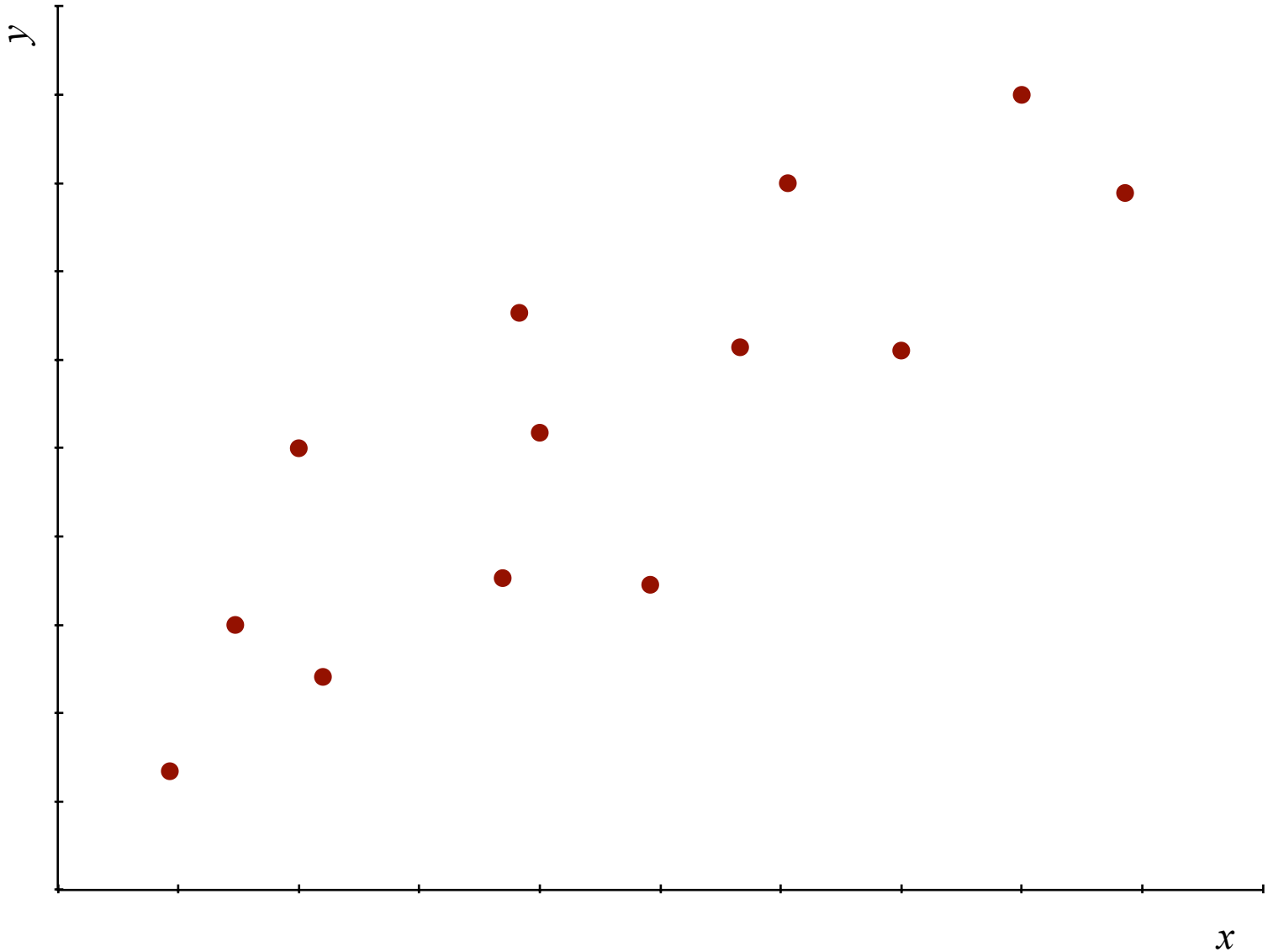
Formally, OLS solves:

$$\min_{\alpha,\beta} \sum_i (\varepsilon_i)^2 \tag{2}$$

where

$$\varepsilon_i \equiv \log(Y_i/L_i) - \alpha - \beta \log(K_i/L_i) - \gamma H_i \tag{3}$$

OLS Illustration



Multiple regression

Typically one adds other "covariates" to a regression (not just H_i).

- The idea is to "hold constant" other things.
- E.g.: schooling, region, democracy

The model is then

$$\log(Y_i/L_i) = \alpha + \beta \log(K_i/L_i) + \gamma_1 X_{i,1} + \cdots + \gamma_J X_{i,J} + \varepsilon_i \quad (4)$$

or in compact notation

$$\log(Y_i/L_i) = \alpha + \beta \log(K_i/L_i) + \sum_j \gamma_j X_{i,j} + \varepsilon_i \quad (5)$$

- X_{ik} is the value of regressor j for country i

Multiple Regression

In (common) matrix notation

$$\underbrace{\log(Y/L)}_{Nx1} = \alpha + \underbrace{\log(K/L)}_{Nx1}\beta + \underbrace{X}_{N\times J}\underbrace{\gamma}_{J\times 1} + \underbrace{\varepsilon}_{J\times 1} \tag{6}$$

Each row is the equation for one observation

$$\log(Y_i/L_i) = \alpha + \log(K_i/L_i)\,\beta + \underbrace{\sum_j X_{i,j}\gamma_j}_{\text{row } i \text{ of } X\gamma} + \varepsilon_i \tag{7}$$

OLS now still finds the values of all **regression coefficients** $(\alpha, \beta, \gamma_j)$ that minimize the sum of squared residuals.

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:

$$Y = \alpha + X\beta + \varepsilon \tag{8}$$

$$= \underset{(0.012)}{0.123} + \underset{(0.45)}{2.34}X + \varepsilon \tag{9}$$

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)

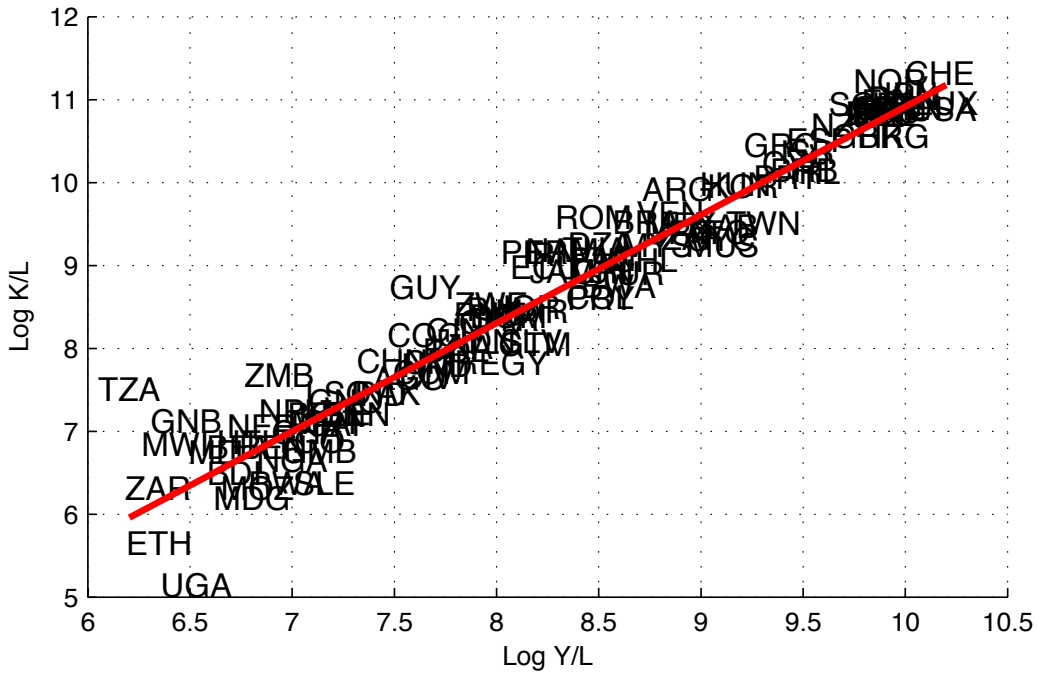
- what does that mean?

R^2: measure of "fit"

- $R^2 = 1 - [\text{residual sum of squares}] / [\text{total sum of squares}]$
- 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.

Suppose this remains true when we add other regressors (the X).

Are we done?

Regression Analysis: Interpretation Issues

Interpreting Regression Results

Suppose we regress

$$\ln(Y/L) = \alpha + \beta \ln(K/L) + X\gamma + \varepsilon \tag{10}$$

and find $\beta = 0.5$

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 cause and effect.

Is there an easy way to prove this?

Regressions Do Not Answer Causality Questions

Proof: I can run the regression in reverse:

$$\log(K_i/L_i) = \hat{\alpha} + \hat{\beta}\log(Y_i/L_i) + X\hat{\gamma} + \hat{\varepsilon}_i \tag{11}$$

Either regression is equally valid.

Implication

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

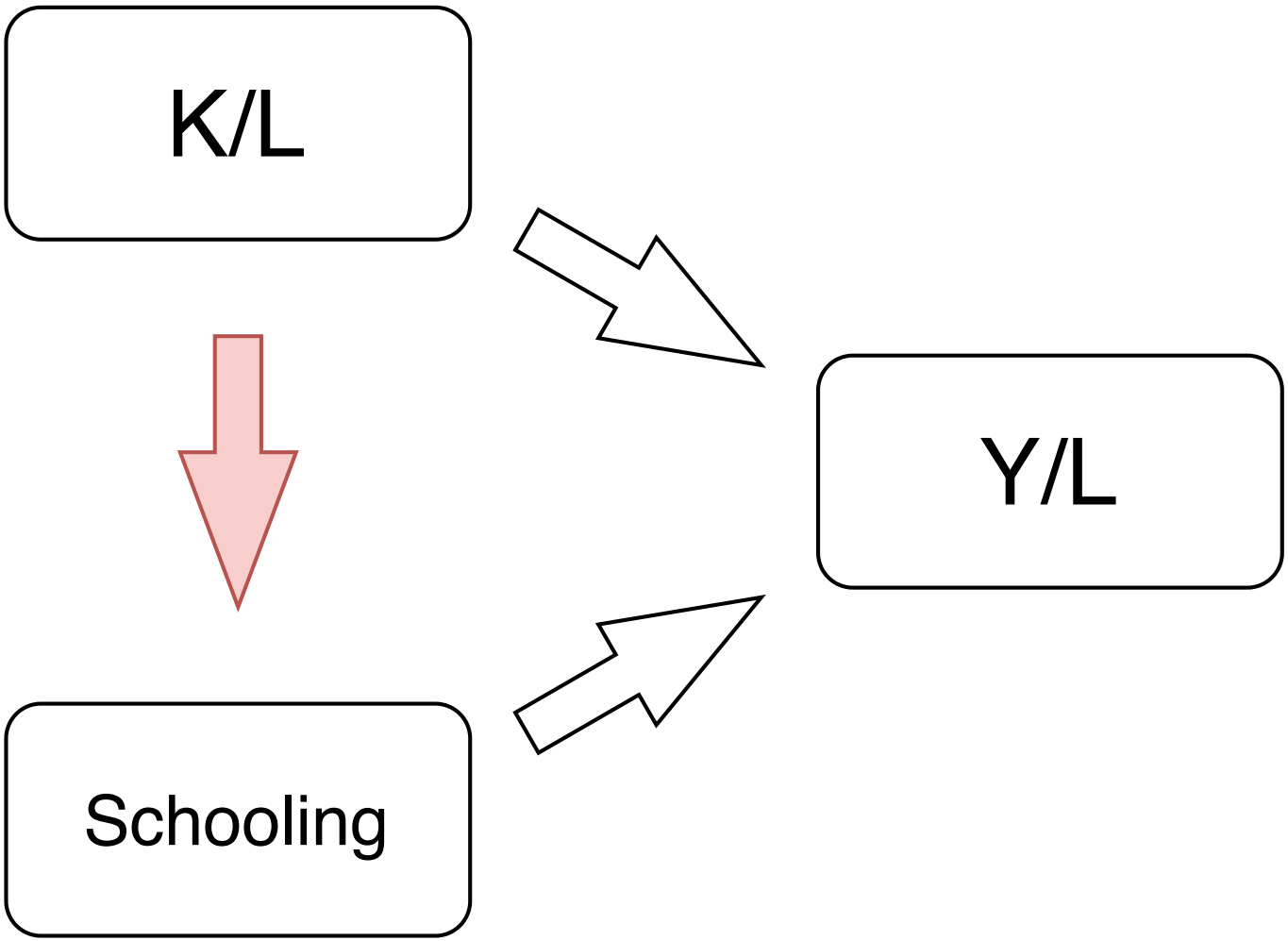
Output depends on capital and schooling

$$\log(Y_i/L_i) = \alpha + \beta_k \log(K_i/L_i) + \beta_s s_i + \varepsilon_i \tag{12}$$

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?

Omitted Variables

A graphical illustration when omitted variables matter:



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?*

Growth regressions

The regression approach has been tried...

I Just Ran Two Million Regressions

Author(s): Xavier X. Sala-I-Martin

Source: *The American Economic Review*, May, 1997, Vol. 87, No. 2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association (May, 1997), pp. 178-183

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/2950909

... and failed.

Interpretation Issues

Fact

No statistical method can answer cause-effect questions.

Two (closely related, partial) exceptions:

- Instrumental Variables (IV)
- natural experiments

Both methods were honored with the 2021 Nobel Prize.

Instrumental Variables

Suppose

$$\log(Y_i/L_i) = \alpha + \beta_k \log(K_i/L_i) + X_i\gamma + \varepsilon_i \tag{13}$$

where we don't know the covariates X.

- either we cannot observe them (example?)
- or we simply don't know the "right" Xs to include

We are looking for the causal effect of K/L on Y/L.

The Idea

Suppose we can find variation in K/L that is

- exogenous (no reverse causality)
- not related to other regressors (X_i) or ε_i

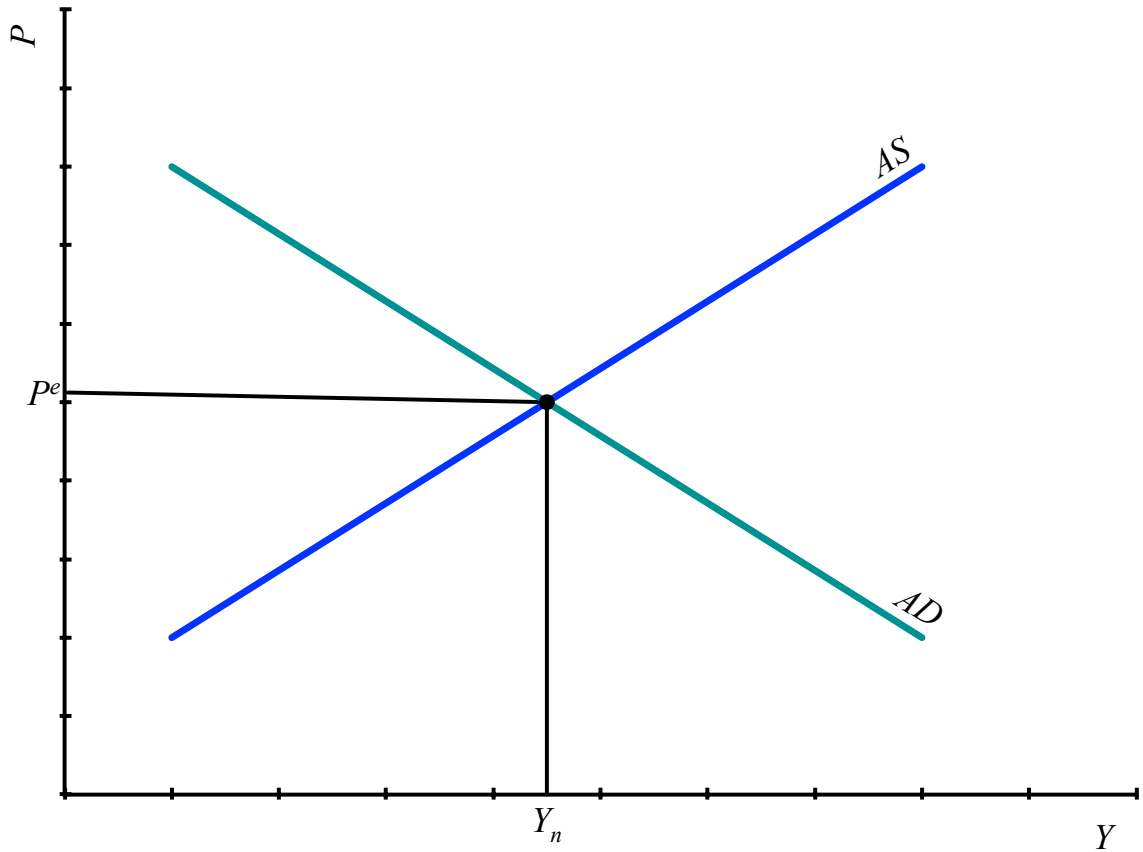
Then we can mimick what a controlled experiment would do:

- isolate this part of the variation in K/L
- see how Y/L varies with it

IV: Classic Example

Suppose we want to estimate the slope of a supply curve.

Why is this hard?



IV: Classic Example

If only AD moves around, estimating AS is easy.

If both curves move around, it's hard.

What if we could identify a variable that only shifts AD, but not AS

- an "instrument"

Then we could find variation in (Y, P) that are related to variation in the instrument.

We could trace out AS.

How To Get the "Right" Variation in X?

The model:

$$\log(Y_i/L_i) = \alpha + \beta_k \log(K_i/L_i) + X_i\gamma + \varepsilon_i \tag{14}$$

The problem:

- K/L is correlated with either X or (worse) ε (omitted vars)

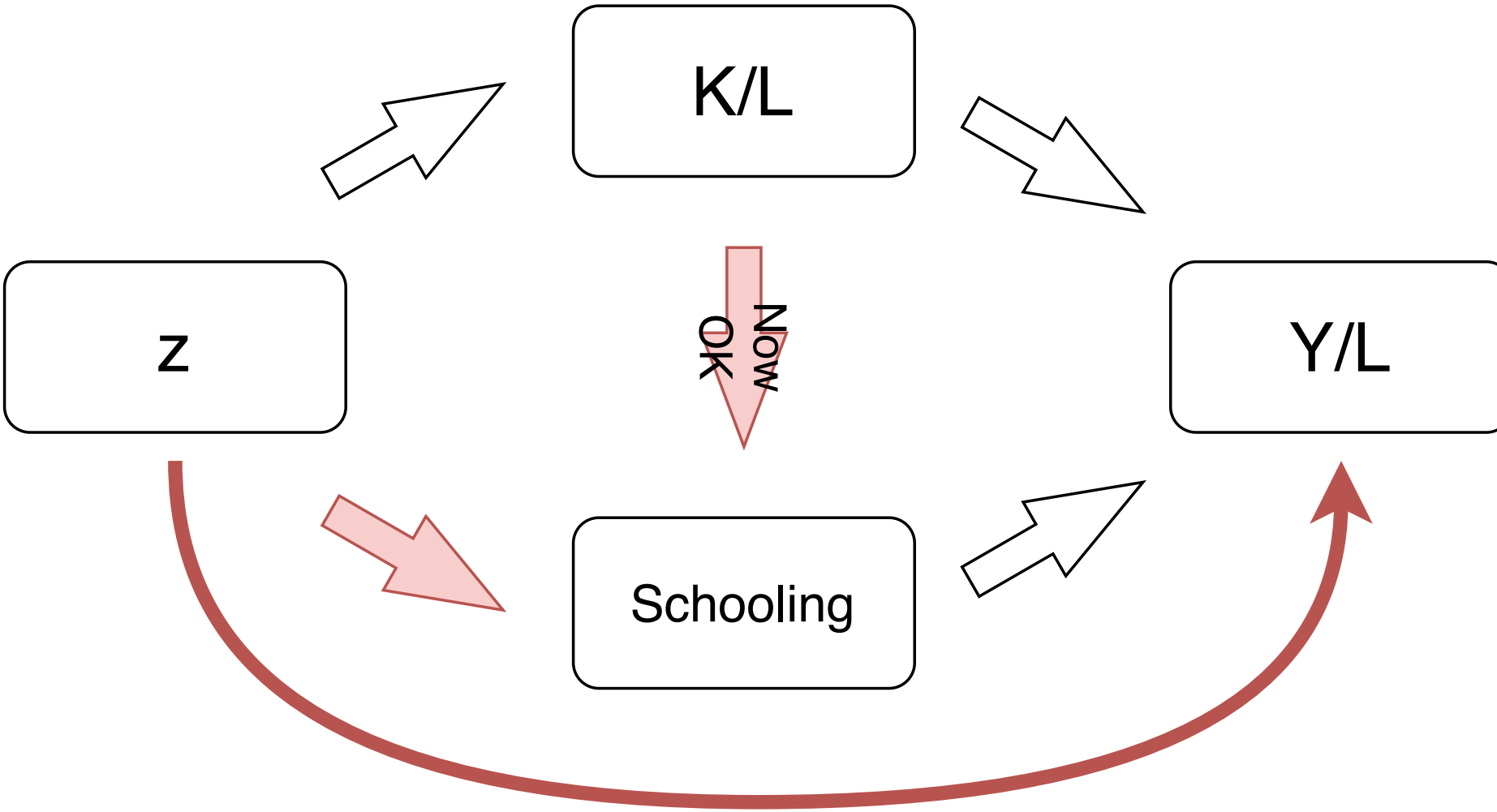
Suppose we also have

$$\log(K_i/L_i) = \delta + \beta_z z_i + \varepsilon_i \tag{15}$$

Exclusion restriction: z has no direct effect on output (it is not part of X)

- this is the key assumption that makes IV "work"

Exclusion Restriction



Instrumental Variables

Then the following works:

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

$$\log(Y_i/L_i) = \alpha + \beta_k \underbrace{\log\left(\widehat{K_i/L_i}\right)}_{\text{predicted}} + \varepsilon_i \tag{16}$$

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

IV intuition

What goes wrong with OLS?

- Omitted variable bias: regressing output on capital gives the wrong coefficient
- because other X are high when capital is high (human capital, institutions, ...)

A Randomized Controlled Trial (RCT) would randomly assign capital to observations.

- Then the capital regressor would be independent of all X
- OLS would work: the average gap between high and low K observations is also the causal effect of K and output.

IV does something similar.

It finds variation in K/L that is not correlated with omitted regressors.

IV Intuition

Key: the exclusion restriction

- 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.

This is the key limitation of IV: it's hard to find instruments.

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

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: this is often the only viable approach.

Reading

Good reference for econometrics (practical issues and interpretation) are:

- Kennedy (2008), Angrist and Pischke (2008), Angrist and Pischke (2014)

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

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

— and — (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

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

Jones, C. I. and Vollrath, D. (2013). Introduction to economic growth. 3. uppl.

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