How to explain things

Suppose you observe variation in something & want to explain it
(e.g. differences in *Y/L*
across countries - development
from year to year in one country - business cycles)

How do you explain it?

1) Form a theory
(e.g. $Y = A K^{\alpha}(HL)^{1-\alpha}$ *where* $\alpha = 1/3$
Keynesian IS/LM or IS/MP model)

2) Derive testable implications
(e.g., if if *K* falls 30%, *Y/L* should fall by 10%
if I increase *r*, *Y/L* falls)

3) Run experiments. “Treatment” (e.g. destroy *K*, increase *r*), observe effect on *Y/L*
but other factors than treatment can affect *Y/L*

What if you can’t reliably control other factors? Perform treatment

- many times on many economies, get a big enough sample for inference

- *randomly* to make sure treatment isn’t correlated with other factors

What if you can’t run experiments?
Here’s what you *shouldn’t* do:

collect data where treatment variable and variable you want to explain wriggle

across time (time-series)
units (cross-section)
time & units (panel)

and observe co-movements between variables.

because other determinants/relevant conditions will be correlated with treatment

1) Anything that changes one economic variable changes others too - general equilibrium!
2) Reverse causation

e.g.

- production function $Y = A K^{\alpha}(HL)^{1-\alpha}$ with treatment ΔK

1) other variables that affect *Y,* perhaps correlated with *K*:

good institutions attract intl. investment (raise K), promote education (raise H)

2) reverse causation from *Y* to *K* if saving/investment is a fraction of *Y*

- IS curve with treatment Δr

1) other variables that affect *Y*, perhaps correlated with *r*: *G, Y* (exports),* Y^{e}

2) reverse causation from Taylor rule $r = \bar{r} + \alpha(Y - \bar{Y}) + \beta(\pi - \pi^{*})$

How to approach nonexperimental questions

1) Ignore reverse causality and correlation with other determinants.

- old papers
- econometrics field, where goal is to demonstrate technique not answer questions

2) Answer question for a model, not real world

In a model, you can exogenously wriggle variable of interest.
To quantify parameters, assume model is true and fit reduced forms to data

3) Identify and control for all possible other determinants. You can't!

But at a *minimum* you must avoid obvious omitted variables bias.

Control for other measurable variables suggested by theories/research
your audience is familiar with.
If you can't include all at once (not enough degrees of freedom),
show your result is "robust" to various sets of possible control variables
e.g. "Extreme bounds analysis"

but this can at best show robust correlation, cannot prove causation

4) Natural experiments

Natural experiments

Find some wriggles in treatment variable that are uncorrelated with other determinants

Often involves “instrumental variable:”

- something measureable that causes or indicates wriggles in treatment variable (“relevant”)
- uncorrelated with other determinants (“valid”)

e.g. in IS curve, exogenous variations in central bank setting of r

In terms of regressions

$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (\varepsilon = \beta_3 X_3 + ...)$

X_1 "Troublesome variable" $\qquad$ X_2 "Control variable" $\qquad$ $X_3 ...$ Unmeasurable determinants

General equilibrium problems: $\quad X_3 = \alpha_{31} X_1 + ... \qquad X_1 = \alpha_{13} X_3 + ... \qquad \begin{array}{l} X_1 = \alpha_{14} X_4 + ... \\ X_3 = \alpha_{34} X_4 + ... \end{array}$

Reverse causality:

$$X_1 = \alpha_{Y1} Y + \alpha_{15} X_5 + ... \quad \rightarrow \quad X_1 = \alpha_{Y1}\left(\beta_1 X_1 + \beta_2 X_2 + \varepsilon\right) + \alpha_{15} X_5 + ...$$

correlation with ε $\nearrow$

Instrumental variables *Z:*

1) Relevant: Z is correlated with X_1

$$X_1 = \gamma_{Z1} Z_1 + \gamma_{Z2} Z_2 + ...$$

2) Valid: Z is *uncorrelated* with ε (not cause or effect or common cause or...)

$$X_3 = 0 \times Z_1 + 0 \times Z_2 + ... \text{ "Exclusion restriction"}$$

Methods

1) 2SLS Good for small samples and/or probably-homoskedastic ε ‘s

1) “First stage” Regress X_1 on Z**.**

2) Use first-stage coefficients to get predicted values $\hat{X}_1$

Note there’s lots of variation in X_1 not captured by $\hat{X}_1$

3) “Second stage” Regress Y on X_1

Note that it wouldn’t be right to use SE’s from second stage to judge significance -
this wouldn’t account for uncertainty about coefficients of first-stage regression.
2SLS SE’s account for this.

2) GMM Good for large samples, heteroskedastic ε ‘s
Can allow for correlations across observations’ residuals.
E.g. “clustering” (common in panels):
ε ‘s correlated across some observations, uncorrelated across others

Problem! if instruments are **“weak”** (relevant but only weakly correlated with X_1)
SE’s *too small* and estimated coefficients *biassed* toward OLS values
(though *consistent* as $n \Rightarrow \infty$)
even in *very large* samples

How do you find instruments?
How do you know if possible instruments are *relevant* and *valid*?

1) Think about why X_1 varies. List all possible reasons.

2) Think about other determinants of Y. List all possible X_3 ‘s.

3) Which reasons for X_1-variation are unrelated to X_3 and *measurable*?

Of course, you must think about what causes Z to vary, and on and on...

This is not a statistical issue. It’s knowledge about the world (including economics).

but there are..

Statistical methods to test whether candidate *Z'*s are relevant, strong & valid

1) Look at "first-stage" regression.

Check sign, significance, magnitude (plausible?) of coefficient on *Z*.
Because of "weak instrument" problem, *t*-stats (or *F*-stats for multiple *Z*'s)
need to be *lots bigger than 2*. "Rule of thumb": at least 10.

2) Look at "reduced form" regression of *Y* on *Z*

$$Y = \beta_1\left(\gamma_{Z1}Z_1 + \gamma_{Z2}Z_2 + \ldots\ldots\right) + \beta_2 X_2 + \varepsilon$$

Sign, magnitudes of coefficients should be consistent with story.

3) If you have more than one *Z* (X_1 is "overidentified") you can test one *Z* against others,
see if results are consistent across *Z*'s.

Validity/exclusion restriction means all *Z's* *uncorrelated* with $(Y - \beta_1 X_1 - \beta_2 X_2)$

a) Get estimates $\hat{\beta}_1, \hat{\beta}_2$
b) Calculate $(Y - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2)$
c) Test hypothesis that *Z's* are uncorrelated with $(Y - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2)$

Sargan test (for 2SLS), Hansen J-test (for GMM)

Z's are OK if you "fail to reject" hypothesis.