Econometrics II: Econometric Modelling

Jürgen Meinecke

Research School of Economics, Australian National University

31 August, 2018

Assignment 1

How did it go?

What have you learned?

- Is globalization good or bad for the environment?
- Race to the bottom or gains from trade?

Logistics for week 7 (after the break)

We will return the assignments in your next tutorial

You can keep your assignment and take it with you

However, if you want to qualify for a remark you need to:

- raise and explain any concerns regarding the marking with your tutor as soon as possible during the week 7 tutorial
- hand your assignment back to your tutor by the end of the week 7 tutorial

Once you leave the tutorial room with your assignment, you cannot as for a remark

Roadmap

Introduction

Instrumental Variables Estimation

Where Do Instruments Come From? (Part 2)

Last week we saw that the hardest part of TSLS estimation is to come up with instruments

Broadly speaking, we need to find explanatory variables that weren't included in the first place and that, at the same time, are correlated with the endogenous variables

A proper IV feels like it is almost randomly assigned

Here are two more examples from applied work in econometrics

Example 3: Effect of Heart Surgery on Longevity

McClellan, M, McNeil B. J., and Newhouse J. P. (1994),
"Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?"
Journal of the American Medical Association

Research question:
Does cardiac catheterization (CC) improve longevity of heart attack patients?

- Y_i = survival time (in days) of heart attack patient
- $X_i = 1$ if patient receives CC, zero otherwise

What is the endogeneity problem here?

What do authors suggest as IV:

Distance to the nearest CC hospital minus distance to the nearest "regular" hospital

Justification:

- If a CC hospital is far away, patient won't be taken there and won't get CC
- If distance to CC hospital doesn't affect survival, other than through effect on CC, then $\text{Corr}(distance, u_i) = 0$, so exogenous
- If patients location is random, then differential distance is "as if" randomly assigned

Findings:

OLS estimates significant and large;

but TSLS small and insignificant

Example 4: School Competition

Hoxby, C. M. (2000), *"Does Competition Among Public Schools Benefit Students and Taxpayers?"* American Economic Review

Research question:

What is the effect of public school competition on student performance?

Data:

- cross-section, US, metropolitan area, late 1990s (n = 316)
- Y = 12th grade test scores
- X = measure of choice among school districts
  (# of school districts per student in metro area)
- W = lots of control variables

What is the endogeneity problem here?

What does author suggest as IV:

Z_i = # small streams/rivers in metro area

Justification:

- many present day school districts originated in the 18th and 19th centuries
- presence of streams/rivers determined school district boundaries back then
- a city with many rivers would have more school districts than an otherwise comparable city (with fewer rivers)
- even today, these two cities tend to have different numbers of school districts
- therefore the number of streams/rivers is positively correlated with the measure of competitiveness of the public school market (IV relevance)
- but is it also exogenous?

Findings:

- student achievement is higher in school districts with more competition
- since author uses TSLS estimation, these results permit to interpret the effect of school competition on student outcomes as causal

Problem Solving Exercises

1. Prove that $s_{\hat{X}Y} = \hat{\pi}_1 s_{ZY}$ and that $s^2_{\hat{X}} = \hat{\pi}^2_1 s^2_Z$.
2. We know that $\hat{\beta}_{1,TSLS} = \frac{s_{\hat{X}Y}}{s^2_{\hat{X}}}$.
   Prove that
   $$\hat{\beta}_{1,TSLS} := \frac{\sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^n (Z_i - \bar{Z})(X_i - \bar{X})} = \frac{s_{ZY}}{s_{ZX}}.$$
3. Prove that the TSLS estimator is consistent.
   Use the regression equation $Y_i = \beta_0 + \beta_1 \hat{X}_i + w_i$ where $w_i := \beta_1 (X_i - \hat{X}_i) + u_i$ and show that the sample covariance between $\hat{X}_i$ and w_i converges to zero (in probability).