

Causal Inference for Policy Evaluation – Spring Semester 2025

Lab Assignment 3

Due date: Friday, 02.05.2024 (at 23:59)

Instructions

Please send your solutions to mario.bernasconi@unibas.ch. The solutions should include two files. One file with your code saved in .R format. A second file with written answers, including tables and graphs, should be sent as a .pdf file and is limited to a maximum of 2 pages (12 points, single spacing). PhD student groups are allowed one more page as they have one more question. Any text or output beyond that will not be considered.

Instrumental Variables (20 points for MSc, 25 for PhD)

1. We want to estimate the effect of fertility on *male* labour supply similarly to Angrist and Evans (1998) using the data set *AngristEvans1980-reduced.RData*. It is a sample of women aged 21 to 35 who were older than 15 at their first birth, and have at least two children with the second child being older than one (these are the restrictions we imposed in the IV lab session).
 - (a) First, restrict the data to married couples. Then, compute the average age and income of mothers and fathers separately when their first child was born. How could what you observe affect the intra-household division of (market) labour supply, domestic chores and childcare? **(1 point)**
 - (b) Estimate the effect of fertility (*morekids*) on men's hours worked per week (*hourswd*) with 2SLS using the same sex instrument (*samesex*) and without any control variable. Report the 2SLS estimate and also the first stage **(1 points)**
 - (c) Now estimate the effect of fertility (*morekids*) on men's hours worked per week (*hourswd*) with 2SLS by using 'two boys' (*boys2*) and 'two girls' (*girls2*) jointly as instru-

ments (and without any control variable). Report the 2SLS estimate and also the first stage. **(1 points)**

(d) Compare and comments on

- the different approaches in points (b) and (c)
- the 2SLS estimates you obtained in points (b) and (c)
- the estimates from the first stages in points (b) and (c)

(3 points)

(e) Now estimate the effect of fertility (*morekids*) on men's hours worked per week (*hourswd*) with OLS. Compare the OLS estimate to the 2SLS you got before and also the respective standard errors. What do you conclude? Which one would you use for inference? **(2 points)**

2. Keep using *AngristEvans1980_reduced.RData*.

(a) Check what is the median age of women. What is the share of women with more than 2 children among those below vs. above the median age? **(1 point)**

(b) Using the restricted sample of women below the median age, provide an OLS estimate of mother's weeks worked (*weeksm*) on more than two children and the following covariates (for the mother): age, age at first birth, black, hispanic, and other race. Why would you control specifically for these variables? Are these control variables sufficient to alleviate potential endogeneity concerns? Why yes or why not? **(2 points)**

(c) Estimate the same equation as in (b) but with 2SLS and using the same sex instrument. Compare your results to the OLS estimates and comment. **(2 points)**

(d) Why might the sample restrictions you imposed (women below the median age) be problematic for the 2SLS approach? **(1 point)**

3. Suppose you have to replicate the study with recent data from Switzerland where

- the fertility rate amounts to roughly 1.4 children per women
- the average age of a mother when their first child was born is around 31 years
- female (male) labour market participation is just above 60% (70%)
- among employed women (men), the share of part-time work is around 60% (20%).

- (a) Independent of the instrument used, Angrist and Evans (1998) find no economically and statistically significant effect on male labour supply. Check the employment rates in *AngristEvans1980_reduced.RData* for men and women, and how common part-time work is among employed men and women. Compare the results with the rates for Switzerland outlined above. Based on that, discuss how you would expect the effect of fertility on male labour supply in Switzerland today to differ from those in Angrist and Evans (1998). **(2 points)**
 - (b) Discuss how you would use a 2SLS approach to estimate the effect of fertility on labour supply in Switzerland today. **(2 points)**
 - (c) Discuss how you would modify the sample restrictions that we imposed in the IV lab session on the data if we were to use recent data from Switzerland. **(2 points)**
4. **This question is for PhD students only.** Angrist and Evans (1998) present two different instruments to estimate the effect of fertility on labour supply. Could you use this fact to test the validity of the instruments?
- (a) Which test can be performed to test for the exogeneity of the instruments in a parametric IV? Explain the logic behind it, and clearly explain what hypothesis we can test. **(3 points)**
 - (b) Under which assumptions will the test be informative about instrument validity? **(2 points)**

Regression discontinuity design (5 points)

6. Reading for the next lab session: Meyersson (2014). You may skip section 6 in the paper.
- (a) In addition to data from the 2000 population census, the author uses information on outcomes in 1990, which is before the assignment to the treatment. What does he use it for and why is it in support of the identification strategy? **(2 points)**
 - (b) The author determines the optimal bandwidth using the algorithm developed by Imbens and Kalyanaraman (2012). Briefly explain why it yields a smaller optimal bandwidth for women. **(1 point)**

- (c) For women, the author proposes and tests a mechanism that can explain the positive RD estimate (despite a negative raw correlation). Explain this mechanism and comment on the test procedures. **(2 points)**