Lab for running an IV

Matrix Preparation (optional)

1.) Starting from file Lab_09.RScript, try to run yourself what you saw in "Matrix_03" the primer session.

OLS regression on foot:

In what follows use the startscript.R provided to OLSqData

- ⇨ Take a look at the data, which variables do you have?
- ⇨ Now use cbind to build 2 matrices:
  - Using DT$const, DT$x1, DT$x2 and DT$x3 to build matrix X, and
  - Using DT$y5 to build matrix y
- ⇨ Now compute X' (transpose of X)
- ⇨ Next compute X'X . Report your result.
- ⇨ Next compute the inverse of X'X. Report it.
- ⇨ Next compute X'y. Report
- ⇨ Lastly multiply the inverse of X'X with X'y. Report your result. What is this?
- ⇨ Finally run a few comparisons.
  - Regress y1 on x1, x2, and x3 in a linear model.
  - Divide XtX by n (number of observations) and compare it to cov(DT), comment.

IV – Exercise

In what follows, use the dataset 'IV_Data', in which the x-variables are endogeneous and Z1, Z2, Z3 and Z4 will be candidates for instrumental variables.

This is the Var-Cov matrix that I used for simulating the data:

Correlations	Const	x1	x2	Z1	Z2	Z3	Z4	eps
Const	1	0	0	0	0	0	0	0
x1	0	1	0.12	0.5	0	0.2	0.01	0.6
x2	0	0.12	1	0.4	0	0	0.001	0.02
Z1	0	0.5	0.4	1	0.6	0	0	0.6
Z2	0	0	0	0	1	0.4	0	0
Z3	0	0.2	0	0	0.4	1	0	0
Z4	0	0.01	0.001	0	0	0	1	0
eps	0	0.6	0.02	0.6	0	0	0	1

Compare the estimates of the full OLS model and the IV Regression. - Why are the estimated coefficients different?

Looking at the Var-Cov Matrix, what bias would you expect for x1 and what bias for x2?, calculate! (Hint: the X – matrix for calculating the bias consists of x1 and x2., i.e. you can disregard the z's and the constant for this calculation. E(X,u) is given by the "eps" - column)

Which of the X Variables is endogenous with y (e.g. a simultaneity problem)?

If you can mend only one of the variables, which one would you tackle and why?

You have 4 candidates that you can use as an instrumental variable for X1, but

⇨ One is itself endogenous.

⇨ One is a weak instrument

⇨ One is irrelevant and

⇨ Only one is valid.

Which is which?

Now try instrumenting for x1!

- Before you go ahead, consider for each z if you can or cannot use it as instrument. Explain, why not, or, if you can, explain what the assumptions would be and explain the exclusion restriction.
- Refer to the "IV-assumptions" on the slides for your argument.
- Independent of your answer in the previous points: consider using $Z4$ as an instrument for X_2:
    - Which of the two assumptions can you test? Is it satisfied?
- Now pick your most preferred instrument or combination of instruments:
    - Do 2SLS, describe which steps you have to take?
    - Write a code that separately runs the first stage and the second stage regression.
    - Rehash the MM-IV Estimator that we saw in class. (provide it as answer)
    - What variables will you include in the Z matrix?
    - Write a code that directly implements the estimator in matrix notation, using either the data.table command or the matrix command.
        - Build a Z-matrix consisting of the instruments you want to use (including all exogeneous x).

        - Build an X-matrix consisting of the X variables
        - Build a y-matrix that just consists of the vector y
    - Now interpret the coefficient estimates that you got
        - In the direct Matrix IV.
        - In the 2SLS.
    - Share any thoughts, questions or confusion on this exercise. (Hint: it's likely that the final estimation result is not satisfactory – maybe you can see why, but don't worry if you cannot.)

Comparing IVregs:

Using the ivreg library and commands, run all ivregs for all 4 instruments separately, and also run it for 2-3 combinations of ivs that you would like to try.

- ⇨ Compare the 4 single-iv results and discuss.
- ⇨ Use the one that works best and compare it to the combinations you picked.
- ⇨ Also compare the ivreg result to your own "on foot regression"

What is your "best estimate" of the coefficients that I actually used?

Reading:

What are the IV-Assumptions that are provided in the Angrist book?

Compare them to the IV-Assumptions on the slides.

- ⇨ Are there more, less, which are the same, and which are different?