EAE6029 - Econometria I
Lista 2

Prof: Pedro Forquesato

Faculdade de Economia, Administração e Contabilidade
Universidade de São Paulo

2023/1

Exercise list 2

Analytical problems

1. Let's analyze the following potential outcomes model: $h(0,u) = \alpha + \gamma_0 W + u$ and $h(1,u) = \alpha + (1-W)\gamma^L + W\gamma^H + u$, where W is a dummy variable (a "type", let's say sex). What is the ATE? And ATT? And if $W = 1$, what are the ATT and ATE? How would you identify γ^H and γ^L using a linear projection? Now consider the linear projection of $Y = h(T, W; u)$ on T and W: $Y = \beta_0 + \beta_1 T + \beta_2 W + e$. What does β_1 identify?

2. Consider a single variable model $Y = \beta X + e$ and $X \in \mathbb{R}$, and consider an IV estimator where $Z \in \{0, 1\}$. Find a simple expression for the estimator in this context.

3. In the usual structural model $Y_1 = Z_1'\beta_1 + Y_2'\beta_2 + e$ and $Y_2 = Z'\Gamma + u$, Γ is $l \times k$, $l \geq k$, it is stated that a condition for identification is $\text{rank}(\Gamma) = k$. Show that this is true.

   Correto: Gamma é l_2 x k_1.

4. Take the linear model $Y = X'\beta + e$, with $\mathbb{E}[e|X] = 0$, with $X \in \mathbb{R}$. Is $Z = (X, X^2)$ a valid instrument for β? How does this differ from OLS?

   Rank(Gamma Barra) = k = k_1 + l_2

5. Consider a supply and demand as below, where Y is income and W wages. Discuss the conditions for its identification:

$$[\text{S:}]\ Q = \alpha_S + \beta_S P + \gamma_1 Y + e_S$$

$$[\text{D:}]\ Q = \alpha_D + \beta_D P + \gamma_2 W + e_D$$

6. In the standard IV structural model $Y = X'\beta + e$ and $X = Z'\Gamma + u$, consider the usual control function $e = u'\eta + v$. Show (algebraically) that $\mathbb{E}[Xv] = 0$.

7. Consider a IV structural model with a real-valued endogenous variable X, where $Y = \alpha + \beta_1 X + \beta_2 X^2 + e$ and $X = \gamma_0 + \gamma_1 Z + u$, with $\mathbb{E}[e|Z] = 0$. Is X^2 endogenous? If

we estimate an IV model with instruments $(1, Z, Z^2)$, under what conditions is the model identified?

8. In the standard structural model $Y_1 = Z_1'\beta_1 + Y_2'\beta_2 + e$ and $Y_2 = Z'\Gamma + u$, show that we can test the null $H_0 : \beta_2 = 0$ only by the reduced form equation $Y_1 = Z_1'\lambda_1 + Z_2'\lambda_2 + v$.

9. Consider two i.i.d. samples, one for men and other for women, and the regression $Y_l = X_l'\beta_l + e_l$, where l is men or women. Think X_l are endogenous, and we have valid instruments Z_l, $l = 1, 2$. Develop a statistic for the test $H_0 : \beta_1 = \beta_2$ and explain briefly how you would estimate it.

Computational/interpretative problems

For this list, we will use the `AK1991.dta` file provided by the textbook author. You should **provide the code** and the results *together*.

```
library(haven)
library(kableExtra)
library(knitr)

# setwd() <- this might help

#  you can download the file directly in R, or just do it manually
# url <- "https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics%20Data.zip"
# download.file(url, "./econ_data.zip")
# unzip("./econ_data.zip")

ak1991 <- read_dta("./AK1991.dta")
knitr::kable(head(ak1991, 10))
```

ageq	edu	logwage	married	state	qob	black	smsa	yob	region
47.00	12	6.245846	1	1	1	1	1	1933	0
46.25	12	5.847161	1	48	4	1	1	1933	0
50.00	12	6.645516	1	2	1	1	1	1930	0
47.00	16	6.706133	1	22	1	1	1	1933	0
42.25	14	6.357876	1	42	4	1	1	1937	0
44.25	12	5.441835	1	5	4	1	1	1935	0
42.00	12	6.390660	1	42	1	1	1	1938	0

ageq	edu	logwage	married	state	qob	black	smsa	yob	region
50.00	12	4.607667	1	48	1	1	1	1930	0
40.50	12	6.553961	1	5	3	1	1	1939	0
43.75	7	7.017041	0	2	2	1	1	1936	0

Here we will take a look at Angrist and Krueger (1991) analysis of returns to education. As previously mentioned, there is a long tradition in economics (and econometrics) trying to identify the **Mincerian equation** of returns to schooling.

AK try to employ an instrumental variables strategy, instrumenting for education by time of birth, since compulsory schooling laws define a minimum age everyone must stay at school (usually 16 or 17).

1. Restrict the sample only to black. Run an OLS regression of log wages on education, urban and married, as well as dummies for year-of-birth, state-of-birth and region-of-residence (these dummies do not need to be reported).
2. Are the OLS estimates trust-worthy? Explain in detail, using examples. If useful, draw a DAG.
3. Now apply the 2SLS estimator, using as instruments quarter-of-birth times year-of-birth (30 instruments). Report the first stage, reduced form and 2SLS estimates.
4. Compare the results of 1 and 2. If we believe in the IV regression, what does it tell us about the OLS bias? Be detailed. If useful, draw a DAG.
5. Do an endogeneity test and an overidentification test. Discuss the results and interpret them precisely.
6. Estimate now the IV regression using the two-stage procedure discussed in class. Are the estimates the same? What about the estimated standard deviation of the estimator?
7. Now estimate the structural model using the control function approach. Again compare the results.
8. Are the instruments strong or weak? Be detailed and precise.
9. A common suggestion in these cases is to lower the number of instruments. Try using only quarter-of-birth (3 instruments). Which would you use in your own research? Be precise why.
10. In class we discussed how instruments identify the local average treatment effect (LATE). How does this affect the interpretation of the results? Explain in detail.
11. Angrist & Krueger (1991) control for region-of-residence and urban. Is there an argument why we might want to avoid these controls? What do you think?