Metrics, PS5

Giselle Labrador Badia

May 2022

Tables with standard errors are provided for all regressions and other relevant analyses that I discuss. All the Julia code pertinent to this assignment is attached.

Part 1: Analytic exercises

Question 1. Assume a returns to schooling model with just 1 unobserved variable ability A distributed $U[0,1]$. Our potential outcomes model for earnings by schooling (treated) or not is

$$Y_1 = 1 + 0.5A$$
$$Y_0 = A$$

Treatment/schooling is determined by

$$D = 1\{-0.5 + A > 0\}$$

(a) Average treatment effect is

$$ATE = \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[1 + 0.5A - A] = \mathbb{E}[1 - 0.5A] = 1 - 0.5\mathbb{E}[A] = 0.75$$

(b) The fraction of the population that takes the treatment is

$$\mathbb{P}[D = 1] = \mathbb{P}[-0.5 + A > 0] = \mathbb{P}[A > 0.5] = 1 - \mathbb{P}[A < 0.5] = 0.5$$

(c) The maximum treatment effect is

$$\max_{A\in[0,1]} TE = \max_{A\in[0,1]} (Y_1 - Y_0) = \max_{A\in[0,1]} (1 - 0.5A) = 0, \qquad \text{attained at } A = 0.$$

The minimum treatment effect is

$$\min_{A\in[0,1]} TE = \min_{A\in[0,1]} (Y_1 - Y_0) = \min_{A\in[0,1]} (1 - 0.5A) = 0.5, \qquad \text{attained at } A = 1.$$

(d) Assume A distributed Normal, $A \sim N(0,1)$. Now the support in unbouded.

The maximum treatment effect is

$$\sup_{A\in[-\infty,\infty]} TE = \infty$$

The minimum treatment effect is

$$\inf_{A\in[-\infty,\infty]} TE = -\infty$$

(e) The average treatment effect on the treated is

$$\begin{aligned} ATET &= \mathbb{E}[Y_1 - Y_0|D = 1] = \mathbb{E}[1 - 0.5A|D = 1] \\ &= \mathbb{E}[1 - 0.5A|A > 0.5] = 1 - 0.5\mathbb{E}[A|A > 0.5] \\ &= 1 - 0.5 * 0.75 = 0.625 \end{aligned}$$

The average treatment effect on the untreated is

$$\begin{aligned} ATEU &= \mathbb{E}[Y_1 - Y_0|D = 0] = \mathbb{E}[1 - 0.5A|D = 0] \\ &= \mathbb{E}[1 - 0.5A|A < 0.5] = 1 - 0.5\mathbb{E}[A|A < 0.5] \\ &= 1 - 0.5 * 0.25 = 0.875 \end{aligned}$$

(f) ATEU > ATET because $TE = 1 - 0.5A$ is decreasing in A; in words, education (treatment) will have a greater effect on the earnings of individuals with low ability. Individuals with high ability will have relatively high earnings with or without education.

(g) The OLS estimand for the effect of D on Y is

$$\begin{aligned} \beta(OLS) &= \mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0] \\ &= \mathbb{E}[1 + 0.5A|A > 0.5] - \mathbb{E}[A|A < 0.5] \\ &= 1 + 0.5 * 0.75 - 0.5 \\ &= 1.125 \end{aligned}$$

(h) The OLS estimand is biased upward for the ATE because treatment is not random but biased towards those with higher ability. More formally, selection into treatment requires a higher ability, thus conditional independence fails.

Question 2. 2) Assume the potential outcomes model with $V = \delta_0 + \delta_1 Z + U_V$, for instrument $Z \in \{0, 1\}$.

(a) Now I prove that this model implies the Angrist and Imbens monotonicity assumption. (Note: I omitted i subscripts, but δ_0 and δ_1 are homogeneous parameters.)

Monotonicity implies that given parameters δ_0, δ_1, $V_i(Z)$ is monotonic in Z for all i.

See that $V_i(Z = 1) - V_i(Z = 0) = \delta_1$ for all i, which means

- $V_i(Z = 1) > V_i(Z = 0)$ if $\delta_1 > 0 \Rightarrow V_i$ is increasing in Z
- $V_i(Z = 1) > V_i(Z = 0)$ if $\delta_1 < 0 \Rightarrow V_i$ is decreasing in Z

- $V_i(Z = 1) > V_i(Z = 0)$ if $\delta_1 = 0 \Rightarrow V_i$ does not change with Z

(b) In this model, a new function for V such that monotonicity does not hold is

$$V_i = \delta_0 + \delta_{i1} Z + U_{iV}, \text{, where } \delta_{i1} \in \{-1, 1\}$$

This model presents heterogeneous treatment effects, now $V_i(Z = 1) - V_i(Z = 0) = \delta_{i1}$ could be 1 or -1. Hence, monotonicity does not hold.

Question 3. Assume U_V is distributed Uniform $[-2, 2]$, and $V = Z + U_V$ with $Z \in \{0, 1\}$. As usual, I assume individuals take treatment when $V > 0$.

(a) The range of U_V values for the complier, defier, always taker, and never taker groups is

- Compliers (C): Take treatment when $Z = 1$ and do not take the treatment when $Z = 0$.

$$\iff V_1 > 0 \text{ and } V_0 < 0 \iff -1 \leq U_V \leq 0$$

- Defiers (D): Take treatment when $Z = 0$ and do not take the treatment when $Z = 1$.

$$\iff V_0 > 0 \text{ and } V_1 < 0 \iff -1 \geq U_V and U_V \leq 0$$

which implies that there are no defiers. From the monotoniity assumption we can also infer that there will not be defiers.

- Always takers (AT): Take treatment when $Z = 1$ and when $Z = 0$.

$$\iff V_1 > 0 \text{ and } V_0 > 0 \iff U_V \geq 0$$

- Never takers (NT): Do not take treatment when $Z = 1$ and when $Z = 0$.

$$\iff V_1 < 0 \text{ and } V_0 < 0 \iff U_V \leq -1$$

(b) Hence, the fraction of the population in each group.

- Compliers (C): $\mathbb{P}(-1 \leq U_V \leq 0) = \frac{1}{4}$
- Defiers (D): $\mathbb{P}(D) = 0$
- Always takers (AT): $\mathbb{P}(U_V \leq -1) = \frac{1}{4}$
- Never takers (NT): $\mathbb{P}(U_V \geq 0) = \frac{1}{2}$

Question 4. Assume there are 2 types in the population. Type 1 has treatment effect $\Delta = 2$ and Type 2 has $\Delta = -1$. 30 percent of the population is Type 1 , 70 percent Type 2. Type 1 s have utility given by $V = Z + U_V$, with $U_V \sim U[-1, 1]$ and Type 2 s have utility given by $V = 2Z + U_V$, with $U_V \sim U[-1, 1]$. Let the instrument $Z \in \{0, 1\}$ and $\Pr(Z = 1) = 0.5$.

(a) ATE is

$$\begin{aligned} ATE = \mathbb{E}[\Delta] &= \mathbb{P}(Type1)\mathbb{E}[\Delta|Type1] + \mathbb{P}(Type2)\mathbb{E}[\Delta|Type2] \\ &= 0.3 * 2 + 0.7 * (-1) = -0.1 \end{aligned}$$

(b) The probability of being treated given the instrument Z is

$$\begin{aligned}\mathbb{P}(D=1|Z=1) &= \mathbb{P}(Type1)\mathbb{P}(U_V>-1)+\mathbb{P}(Type2)\mathbb{P}(U_V>-2)\\ &= 0.3*1+0.7*1=1\end{aligned}$$

$$\begin{aligned}\mathbb{P}(D=1|Z=0) &= \mathbb{P}(Type1)\mathbb{P}(U_V>0)+\mathbb{P}(Type2)\mathbb{P}(U_V>0)\\ &= 0.3*\frac{1}{2}+0.7*\frac{1}{2}=\frac{1}{2}\end{aligned}$$

(c) LATE is same as ATE. See that Z and U_V doe not affect the treatment effect of any of the two groups. Compliers are $U_V \leq 0$ for the two types. More formally:

$$\begin{aligned}LATE &= \mathbb{E}[\Delta|U_V<0] = \mathbb{P}(Type1)\mathbb{E}[\Delta|Type1, U_V<0]+\mathbb{P}(Type2)\mathbb{E}[\Delta|Type2, U_V<0]\\ &= \mathbb{P}(Type1)\mathbb{E}[\Delta|Type1]+\mathbb{P}(Type2)\mathbb{E}[\Delta|Type2]\\ &= ATE=-0.1\end{aligned}$$

Part 2: Monte Carlo exercises

Question 1. Assume log hourly earnings for individual i takes this form:

$$\ln w_i = 1+0.05s_i+0.1a_i+\epsilon_i$$

where s_i is units of observed schooling and a_i is unobserved ability. $\epsilon_i \sim N(0,0.5).a_i \sim N(0,4)$ Units of schooling for each individual i are

$$s_i = 3a_i+z_{i1}+z_{i2}+\eta_i$$

where $\eta_i \sim N(0,1).z_{i1}$ and z_{i2} reflect the cost of schooling for individual i. $z_{i1} \sim N(0,0.1).z_{i2} \sim N(0,25)$. There is another variable z_{3i} which is distributed Uniform $[0,1].z_{1i}, z_{2i}, z_{3i}$ are all independent of each other, and independent of η_i, a_i, ϵ_i.

Solution: The code that simulates data is in $Model1.jl$. For question b) you can refer to table 1 in the table below.

The instrument z_1 affects wages, so it is a relevant instrument. , though has a very small variance so it is a weak instrument. The instrument z_2 is relevant and it is a strong instrument with a lot of variation. The instrument z_3 is not relevant, it does not have any impact on wages. All instruments are exogenous, that is, if anything, they do not affect wages through any variable other than schooling.

The schooling coefficient, β_2 is biased upwards, but it is pretty close taking into account that there is endogeneity. The explanation above about the instruments explain many of the results we see in the table. The coefficient of the iv with the first instrument is not significant which is expected, and the coefficient is very far from the true effect. Surprisingly, the results with z_3 are not as far from the real parameter (half) and it is significant. Anything that involves the valid instrument z_2 is a good estimate. Looking at the F statistic, we see that the strong set of instruments is those that include z_2. The highest F-statistics is attained with only the second instrument. Similar results but with better estimates are obtained by repeating the exercise with more observations.

	OLS	IV1	IV2	IV3	IV12	IV13	IV23	IV123
N=2000								
	0.0001915	39.6	0.0002082	0.0005187	0.000208	0.000521	0.0002082	0.000208
intercept	0.9778	0.8368	0.9773	0.9758	0.9773	0.9758	0.9773	0.9773
	2.705e-7	8644.0	3.698e-7	0.0006591	3.691e-7	0.000662	3.697e-7	3.691e-7
schooling	0.05616	-2.07	0.04932	0.02591	0.04936	0.02581	0.04932	0.04936
F stat	-	0.0005234	7769.0	2.213	3901.0	1.106	3883.0	2599.0
N=50000								
	7.608e-7	7.96e-7	8.204e-7	4.381e-5	8.204e-7	8.022e-7	8.204e-7	8.204e-7
intercept	1.001	1.001	1.0	1.002	1.0	1.001	1.0	1.0
	9.904e-10	3.996e-5	1.316e-9	0.05244	1.316e-9	3.99e-5	1.316e-9	1.316e-9
schooling	0.05624	0.05871	0.05001	0.1335	0.05001	0.05942	0.05001	0.05001
F stat	-	12.55	2.155e6	0.1233	1.078e6	6.332	1.077e6	718400.0

Question 2.

Consider the following potential outcomes model:

$$
\begin{aligned}
Y_0 &= 1 + U_0 \\
Y_1 &= 4 + U_1 \\
V &= -1 + 2Z + U_V
\end{aligned}
$$

with $(U_0, U_1, U_V) \sim N(0, \Sigma)$, and Σ corresponding to the following variancecovariance elements: $V(U_0) = 1, V(U_1) = 1, V(U_V) = 1, \operatorname{Cov}(U_0, U_1) = 0.5$, $\operatorname{Cov}(U_0, U_V) = 0.3$, and $\operatorname{Cov}(U_1, U_V) = 0.7$. Z is a valid instrument with $Z \in \{0, 1\}$, and $\Pr(Z = 1) = 0.3$.

Solution: The code that simulates data is in $Model2.jl$ (a).

Below I show a table with the solutions to questions (b), (c), and (d).

	b)	d)
ATE	2.951	2.951
ATET	3.176	3.131
ATEU	2.826	2.831
OLS	3.592	3.434
ITT	1.979	2.356
IV	2.950	2.915
% compliers c)	0.69	0.82