Assignment 4 - Econometrics II

Floris Holstege, Stanislav Avdeev

Problem 1

We use the allocation to the two judges as an instrument. This instrument is relevant, since one judge (Jones) will sentence more than the other (Smith). But it is also exogenous, since the defendants are randomly allocated to a judge.

I)

In order to assess the effect of a prison sentence on future arrests, we use the Wald-estimator, defined as follows:

$$\beta_{wald} = \frac{P(Y=1|Z=1) - P(Y=1|Z=0)}{P(D=1|Z=1) - P(D=1|Z=0)}$$

Where $P(Y=1)$ is the probability of an arrest in the subsequent 3 years, Z is our instrumental variable, with $Z=1$ being judged by Jones, and $Z=0$ being judged by Smith. $P(D=1)$ is the probability of a prison sentence.

To estimate the probabilities, we simply use the % of each event occuring in our sample. The Wald estimator then becomes:

$$\beta_{wald} = \frac{0.46 - 0.38}{0.7 - 0.4} = \frac{0.08}{0.3} = 26\frac{2}{3}$$

II)

The result implies that if one receives a prison sentence, one is $26\frac{2}{3}\%$ more likely to be arrested in the subsequent 3 years. This effect holds for all compliers.

Since we are working with an instrument, the definition of compliers and always takers changes. Instead of related to the prison treatment, it is not related to the instrument that we use to estimate the actual effect. That is because non-compliance is now only relevant insofar as it effects the observed outcome from this instrument. The definition here is therefore:

$$Z(1) = 1, Z(0) = 0$$

In this case, those are people who under Jones went to prison, but would not under Smith, and those that under Smith would go to prison, but not under Jones. $(0.7 \cdot 0.6) + (0.4 \cdot 0.3) = 0.54$.

III)

The definition of always takers is in this case

$$Z(1) = Z(0) = 1$$

More specifically, an always taker is someone who is sent to prison by both judges. This fraction equals $0.7 \cdot 0.4 = 0.28$.

Problem 2

I)

We use the following formula:

$$n = (\frac{(1.96 + 0.524)}{MDE})^2 * \frac{\sigma^2}{p \cdot (1-p)} = 617$$

Where n is the required sample size, σ^2 is the variance, and p is the proportion of the control group in the experiment.

II)

The sample size has to increase in order to make up for the non-compliance. We calculate this as follows:

$$\text{new n} = n \cdot \frac{1}{0.8^2} = 964$$

Problem 3

I)

To compute the variance of the flu incidence in the control group, we use the variance of a Bernoulli distribution:

$$Var(X) = p(1-p) = 0.62(1-0.62) = 0.24$$

To estimate the number of children that have to participate in the experiment in order to have an effect of 5%, one has to use the following formula for 80% power:

$$n = (\frac{(1.96+0.84)}{0.05})^2 * \frac{0.24}{0.8 \cdot (1-0.8)} = 4615$$

II

As there was a shortage of flu shots, we have a problem of non-compliance. To correctly estimate the size of participants we have to calculate the fraction of the children in the treatment who received a flu shot (66.8%) and plug it in the following formula:

$$\text{new n} = n \cdot \frac{1}{0.668^2} = 10344$$

If there is a problem of non-compliance, one has to invite more participant in the experiment to find a desirable effect.

III)

Table 1 shows the summary statistics for variables that were measured prior to the experiment. There are no statistically significant differences between control and initially assigned treatment group suggesting the randomization was successful. However, there are statistically significant differences between untreated and treated treatment groups for all variables except for a gender of a child. These results suggest that there was a non random distribution of flu shots within the treament group, i.e. children who lived in rural areas where it was harder to deliver flu shots had less probability to receive one.

Table 1: Mean values for observed characteristics by groups

	type_group	boy	age_mother	edu_mother	married	nationality	income	share
1	Control	0.51	26.09	12.34	0.96	0.28	2269.88	0.20
2	Treatment group	0.50	26.03	12.29	0.96	0.27	2286.49	0.80
2.1	Treated treatment	0.50	26.59	12.52	0.98	0.24	2373.87	0.54
2.2	Untreated treatment	0.50	24.88	11.83	0.94	0.34	2110.71	0.27

IV)

First we estimate the model with only a dummy variable "Flu Shot" and then subsequently include other variables so our extended model is the following:

$$Flu_i = \alpha + \delta FluShot_i + GenderChild_i + AgeMother_i + EducationMother_i + Married_i + Nationality_i + Income_i + U_i$$

Note that we have to use robust stadard errors as we have a binary independent varaible. Table 2 shows the results of the OLS regression. The variable of interest is Treatment and it shows that getting a flu shot decreases the chances to get flu by 27.4% in a base model and it is statistically significant at the 1% level. Including control variables changes the magnitude but not the significance level of the Treatment variable, i.e. an effect of getting a flu shot is 16.3%. In line with the results from the descriptive statistics table we conlude that within the treatment group the flu shots were not randomly delivered to children. Hence, we conlude that instrumental variable is preferred to estimate a causal effect of getting a flu shot on the probability to get flu.

Table 2: The effect of flu vaccine on the probability to get flu in the treatment group

	Dependent variable:	
	Flu	
	(1)	(2)
Treatment	-0.274^{***}	-0.163^{***}
	(0.010)	(0.010)
GenderChild		0.016^{*}
		(0.009)
AgeMother		-0.048^{***}
		(0.002)
EducationMother		-0.030^{***}
		(0.003)
Married		-0.025
		(0.022)
Nationality		0.092^{***}
		(0.010)
Income		0.00001
		(0.00000)
Constant	0.675^{***}	2.185^{***}
	(0.008)	(0.038)
Observations	10,089	10,089
Adjusted R^2	0.067	0.190
Note:	*p<0.1; **p<0.05; ***p<0.01	

V)

We have found that there was a non random distribution of flu shots within the treatment group, thus, we need to use IV estimator. As an instrumental variable for getting a vaccine we use whether or not a child was assigned to the treatment group. The results of IV estimators are in the Table 3. The effect of flu vaccine on the probability to get flu is 19.8% and 19.3% with and without control variables and the effects

are statistically significant at the 1% level. Including control variables do not change the size of the effect suggesting that using IV estimator is preferred in this context.

Table 3: The effect of flu vaccine on the probability to get flu using an instrumental variable

	Dependent variable:	
	Flu	
	(1)	(2)
Treatment	-0.193^{***}	-0.198^{***}
	(0.016)	(0.015)
GenderChild		0.013^{*}
		(0.008)
AgeMother		-0.046^{***}
		(0.002)
EducationMother		-0.027^{***}
		(0.003)
Married		-0.028
		(0.019)
Nationality		0.090^{***}
		(0.009)
Income		0.00000
		(0.00000)
Constant	0.621^{***}	2.153^{***}
	(0.010)	(0.034)
Observations	12,583	12,583
Adjusted R^2	0.059	0.184
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 4 shows the results of the first-stage regression of being assignment to the treatment group on a dummy to get a flu shot. We can see there is a high correlation between these two variables and F-test is greater than 104.7 (Lee, 2020).

Table 4: The first-stage regression of the IV estimator

	Dependent variable:
	Treatment
TreatGroup	0.668^{***}
	(0.009)
Constant	0.000
	(0.008)
Observations	12,583
Adjusted R^2	0.285
F Statistic	$5{,}016^{***}$ (df = 1; 12581)
Note:	*p<0.1; **p<0.05; ***p<0.01

VII)

We assume two things; first, there are no always takers, e.g. people who took the vaccine if they were not assigned to the treatment group. Second, we don't have defiers in our sample when estimating the LATE, e.g. the people who were assigned the treatment but did not comply. We then are left with compliers and never takers (e.g. those who would have not taken the vaccine irrespective of the assigned group).\ \ We define p_{1c} as the percentage of compliers in the treatment group, and p_{0c} the percentage of compliers in the control group.

We define $Y_{i,c}$ as the dependent variable for compliers, and $Y_{i,nt}$ for the never takers. \

$$E[Y|Z=1] = p_{1,c} \cdot Y_{1,c} + (1 - p_{1,c}) \cdot Y_{1,nt}$$
$$E[Y|Z=0] = p_{0,c} \cdot Y_{0,c} + (1 - p_{0,c}) \cdot Y_{0,nt}$$

In order to find the LATE, we define the Wald estimator is defined as:

$$\begin{aligned}\delta_{Wald} &= \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \\ &= \frac{p_{1c} \cdot Y_{1,c} + (1 - p_{1,c}) \cdot Y_{1,nt} - p_{0,c} \cdot Y_{0,c} + (1 - p_{0,c}) \cdot Y_{0,nt}}{E[D|Z=1] - E[D|Z=0]}\end{aligned}$$

We assume $p_{1c} = p_{0c} = p_c$, e.g. the proportion of people who comply in the treatment group is similar to the control group - this is strong assumption in this case, since it seems more plausible that people would divert from the receiving vaccination (due to beliefs about the possible side effects) than the control. We can rewrite the wald estimator as:

$$\delta_{Wald} = \frac{p \cdot (Y_{1,c} - Y_{0,c}) + (1 - p) \cdot (Y_{1,nt} - Y_{0,nt})}{E[D|Z=1] - E[D|Z=0]}$$

$E[D|Z=1] = p_{1,c} = p$, since its just the percentage of people that undertake take the treatment compliers. $E[D|Z=0] = 0$, since there are no people who undertake the treatment who were not assigned (e.g. no always takers). We can thus rewrite the wald estimator as:

$$\begin{aligned}\delta_{Wald} &= \frac{p \cdot (Y_{1,c} - Y_{0,c}) + (1 - p) \cdot (Y_{1,nt} - Y_{0,nt})}{p} \\ &= Y_{1,c} - Y_{0,c} = ATET\end{aligned}$$

where the $ATET$ is the of the average treatment effect on the treated, because we assume once treated everyone complies.

Bibliography

Lee, D. L., McCrary, J., Moreira, M. J., & Porter, J. (2020). Valid t-ratio Inference for IV. arXiv preprint arXiv:2010.05058.