Microeconometrics II

Homework I

Professor: André Portela Student: Rafael F. Bressan

2020-10-27

Part 1: Instrumental variables

You have been provided with a sample of first-born children aged between 10 and 18 years old living with both of their parents. The sample was constructed from Census 2010 . Each entry in the dataset corresponds to a first-born child from both spouses. There can be multiple first-born children in a given family/couple if the first birth from both spouses was a multiple birth. A detailed data dictionary can be found in `dicionario.xlsx`. Your goal is to estimate the causal impact of the number of children in the family (variable `family_number_children`) on the years of education of the couple's first-born child(ren) (variable `first_child_years_of_education`)

Question 1 In this item, we will explore the identifying power of the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions.

*(a) State the MTR and MTS assumptions - **in the most plausible direction** - in this context. Does economic theory have any predictions on the validity of these assumptions? Hint: See the discussion in Ponczek and Souza (2012) .*

Both MTR and MTS definitions are made in Manski and Pepper (2000) , they respectively are:

MTR: Let T be an ordered set. For each $j \in J$,

$w_2 \geq w_1 \implies y_j(w_2) \geq y_j(w_1)$.

MTS: Let T be an ordered set. For each $w \in W$,

$u_2 \geq u_1 \implies E[y(w)|z = u_2] \geq E[y(w)|z = u_1]$.

where W denotes the treatment status, y_j is the potential outcome for person $j \in J$.

In this context, the MTR assumption is most plausible when the number of children in the family **reduces** years of education of the first born child. That is:

$$\text{MTR: } w_2 \geq w_1 \implies y_j(w_2) \leq y_j(w_1) \tag{1}$$

considering that w_i is the number of children in the family and the outcome y_j is years of education of the first born.

In a similar way, the MTS assumption in this context will state that potential outcomes for first born children in bigger families are **lower** than the ones in smaller families.

$$\text{MTS: } w_2 \geq w_1 \implies E[y(w)|W = w_2] \leq E[y(w)|W = w_1] \tag{2}$$

Economic theory, as presented in Ponczek and Souza (2012) Introduction section, predicts that larger families, with a higher number of children will have to share its scarce resources among all children in household, thus leaving the first born with a lower level of education when compared to smaller families. Although, the

empirical evidence is mixed, many of the literature for developed economies show no effect of family size on child quality while for developing countries, the negative hypothesized effect is found.

(b) Report a table with average years of schooling by number of children, as well as the frequency of each value of variable "number of children" in the sample. You may want to use individual sample weights (variable person_weight) in estimation in order to better account for the population of interest.

The total population represented by this sample is 8,274,695, considering a person's sample weight.

Table 1: Years of schooling and number of children by family size.

Children	Years Educ.	Effective Obs.
1	5.87	2281111.58
2	5.91	3547212.08
3	5.63	1587638.98
4	5.06	520484.33
5	4.59	199463.71
6	4.28	83861.10
7	4.08	34215.65
8	3.85	13763.26
9	3.64	4533.67
10	3.60	1624.90
11	3.93	525.70
12	4.39	150.14
13	3.85	80.26
14	4.50	17.29
15	4.00	7.38
16	3.00	4.96

(c) Compute the upper and lower bounds on the ATE of increasing the number of children in the family from 1 to 2,2 to 3,3 to 4,4 to 5,5 to 6 etc. Compute 95% confidence intervals to these bounds using the bootstrap. You should draw samples of households with replacement in order to properly account for the sampling process. You may want to set the probability of sampling a household proportional to the household weight (variable `household_weight`*) in order to better replicate the sampling process.*

First we will derive some results **for binary** treatment, following the MTR and MTS specifications from equations (1) and (2) to get an intuition of what changes from the usual MTR and MTS assumptions. When we impose the MTR assumption in (1), we get a zero **upper bound** on ATE, while the MTS assumption yields a **lower bound** that is equal to the naive difference of means between the two groups. We see that upper and lower bound were interchanged in this new specification. Putting the two assumptions together we have that:

$$E[y_j|W = \bar{w}_n] - E[y_j|W = w_n] \leq \text{ATE} \leq 0 \tag{3}$$

where the treatment $\bar{w}_n$ refers to a family with **more** than n children, and w_n is a family with n children or less, such that $P(W = \bar{w}_n) = \pi$ and $P(W = w_n) = 1 - \pi$.

We begin with the proof that MTR assumption implies the zero upper bound.

Proposition 1. *Given the MTR assumption on* (1) *the upper bound – UB – on the average treatment effect – ATE – is zero.*

Proof. Suppose the treatment is binary with two levels, $w_2 \geq w_1$. The probability of being assigned to w_2 is π. Then the MTR assumption in (1) implies that $y_j(w_2) \leq y_j(w_1)$, and we have the following two inequalities,

$E[y_j(w_2)|w_1] \leq E[y_j|w_1]$ and $E[y_j|w_2] \leq E[y_j(w_1)|w_2]$

The ATE has the following observational-counterfactual decomposition,

$E[y_j(w_2) - y_j(w1)] = \pi E[y_j|w_2] - (1-\pi)E[y_j|w_1] - \pi E[y_j(w_1)|w_2] + (1-\pi)E[y_j(w_2)|w_1]$

Making use of the inequalities in the observational-counterfactual decomposition to obtain an upper bound for the ATE we have that:

$$\begin{aligned} \text{ATE} = E[y_j(w_2) - y_j(w1)] &\leq \pi E[y_j|w_2] - (1-\pi)E[y_j|w_1] - \pi E[y_j|w_2] + (1-\pi)E[y_j|w_1] \\ &= 0 \end{aligned}$$

□

Now we prove that the MTS assumption implies a lower bound on the ATE equal to $E[y_j|w_2] - E[y_j|w_1]$.

Proposition 2. *Given the MTS assumption on* (2) *the lower bound – LB – on the average treatment effect is equal to the difference of means between the treatment groups, ATE* $\geq E[y_j|w_2] - E[y_j|w_1]$.

Proof. Suppose once again the treatment is binary with two levels, $w_2 \geq w_1$. The probability of being assigned to w_2 is π. Then the MTS assumption in (2) implies that potential outcomes for treatment group at $W = w_2$ are lower than in group with $W = w_1$, and we have the two inequalities,

$E[y_j(w_1)|w_2] \leq E[y_j|w_1]$ and $E[y_j|w_2] \leq E[y_j(w_2)|w_1]$

by the observational-counterfactual decomposition the ATE has a lower bound given by:

$$\begin{aligned} \text{ATE} = E[y_j(w_2) - y_j(w1)] &\geq \pi E[y_j|w_2] - (1-\pi)E[y_j|w_1] - \pi E[y_j|w_1] + (1-\pi)E[y_j|w_2] \\ &= E[y_j|w_2] - E[y_j|w_1] \end{aligned}$$

□

Therefore, for a multi-level treatment we can expect the same phenomenon to occur, the upper and lower bound computations will be flipped. This in turns defines that our **upper bound** on ATE will be zero, while the lower bound can be computed by eq. 9.19 from Manski (2009).

$$\Delta(s,t) \leq \sum_{t'>t} E(y|w=t')P(w=t') + E(y|w=t)P(w \leq t) - \sum_{s'<s} E(y|w=s')P(w=s') - E(y|w=s)P(w \geq s) \tag{4}$$

Bellow we present a table with empirical mean of years of study and the distribution number of children across families, a la Table I in Manski and Pepper (2000).

Table 2: Mean of years of study by number of children

w	E[y|w]	P(w)	Size
1	5.8699	0.2718	253592
2	5.9099	0.4172	389234
3	5.6252	0.1960	182834
4	5.0620	0.0677	63182
5	4.5914	0.0271	25249
6	4.2847	0.0119	11059
7	4.0771	0.0050	4694
8	3.8519	0.0021	1971
9	3.6444	0.0008	703
10	3.6032	0.0003	247
11	3.9271	0.0001	96
12	4.3929	0.0000	28
13	3.8462	0.0000	13
14	4.5000	0.0000	2
15	4.0000	0.0000	2
16	3.0000	0.0000	1

In table 3 we have the estimated lower bound for the causal effect and the 5% bootstrap quantile, as in Manski and Pepper (2000) Table II.

Table 3: Lower bound on years of study for first child.

		Lower Bound on Delta(s,t)	
s	t	Estimate	5% quantile
1	2	-0.1443	-0.1052
2	3	-0.3696	-0.3412
3	4	-0.7796	-0.7709
4	5	-1.1631	-1.1647
5	6	-1.4414	-1.4879

Question 2 In this item, we will use an instrumental-variable approach in estimating the causal impact of the number of children on years of schooling of the first child. We will follow the approach in Ponczek and Souza (2012), whereby we first restrict our sample to families with two or more births from the couple (variable `family_number_births` ≥ 2). We then propose to instrument the number of children with `second_birth_ismultiplebirth`, which indicates whether the second birth of the couple was a multiple birth [3]

(a) Make the sample restrictions previously discussed. How many second births are multiple births?

When considering only the families with more than one birth we end up with 644453 observations and the number of second births that are multiple is **9446**.

(a) Under which assumptions does an instrumental variable approach identify a treatment effect in our setting? What treatment effect? Do these assumptions seem plausible to you? Why? Looking at the dataset, do you think any controls should be included? Why?

First we need the usual IV assumptions of relevance and exclusion. Then, for a causal interpretation of the

estimated parameter we need that the instrument must be as good as randomly assigned after controlling for relevant variables. Suppose we have an instrument Z for treatment W, with $Z_i, W_i \in \{0, 1\}$ then:

i) Relevance: $E[Z_i W_i] \neq 0$

ii) Exclusion: $E[Z_i \varepsilon_i] = 0$

iii) As good as randomly assigned: $Y_i(1), Y_i(0) \perp Z_i | \mathbf{X_i}$

In the present case, the instrument variable of choice is a multiple birth for the second born child. Relevance is satisfied since a multiple birth is clearly related to the family size. The exclusion restriction of multiple births has been debated in the literature, but many recognized authors have chosen this instrument on the assumption that having twins is largely random, thus, not related to any possible omitted variable that is related to the outcome years of education of first born child. Finally, from the previous discussion about the exclusion restriction, we do believe this instrument is as good as randomly assigned.

Only if the instrument Z_i were **perfectly** correlated to the the treatment W_i we would be able to identify the ATE. Since this is not usually the case, the instrument is only moderately correlated (i.e. relevant) to the treatment, ATE is not identifiable from this setting. This is the case of imperfect compliance, and in order to identify some other effect from the data we need a fourth assumption:

iv) Monotonicity: $W_i(1) \geq W_i(0)$

That is, the effect of the instrument on the treatment points toward the same direction for every unit. If we assume monotonicity, then the Local Average Treatment Effect – LATE – is identified. The LATE has the meaning of the average treatment effect on the subpopulation of **compliers** only. The compliers are those individuals who would take the treatment if induced to do so (by the instrument variable), but otherwise would refrain.

Actually, the twins instrument is special in the sense it does not allow a subpopulation called **never-takers**. If a multiple second birth is present, there is no way this family ends up with only two children, it has to be three or more. In this special case, LATE coincides with the average treatment effect on the untreated – ATU, see Angrist and Pischke (2008), section 4.4.2 for a discussion.

Therefore, for the instrumental variable `second_birth_ismultiplebirth` we do believe all assumptions are plausible and the LATE can be estimated from data.

The set of control variables could potentially be empty, since our instrument is as good as randomly assigned. If the researcher chooses to control for some characteristics, it is more related to improve the estimate precision than satisfying assumption iii) above.

We do not necessarily think any control **should** be include, although, if we can control for other characteristics that helps in determine the outcome, this inclusion would improve our estimate by reducing the estimator's variance. Such variables could be, age of the first child, whether the first child is a girl. The indicative of first birth is multiple is also a good candidate for control variable, since it may explain the schooling of the multiple first born.

(b) Estimate the treatment effect using 2SLS. Include any covariates you regarded as necessary in the previous item. Cluster your standard errors at the household level (you may also want to weight observations by the person weight). Is the instrument relevant? Why? Comment on your results.

We estimate two models. The first one is a single regression model of years of education on number of children, instrumented by multiple second birth. The second model adds the covariates child age, girl and multiple first birth.

	Model 1	Model 2
(Intercept)	6.662	-3.505
	(0.099)	(0.074)
family_number_children	-0.349	-0.397
	(0.037)	(0.032)
first_child_age		0.714
		(0.002)
first_child_is_girl		0.615
		(0.007)
first_birth_ismultiplebirth		-0.209
		(0.049)

Note: Cluster robust standard errors reported in parentheses. Cluster variable is 'id_household'.

In order to assess the instrument's relevance we report the first stage **F statistic**. For model 1 and model 2 this statistic is, respectively, 5917 and 6143, showing that the chosen instrument is relevant.

The inclusion of control variables didn't change significantly the parameter estimate, as expected. Since our instrument is valid, relevant and randomly assigned, controlling for other relevant variables just improves precision, but do not alter the point estimate. We notice however, that all three control variables are significant, but robust standard error for model 2 improved only slightly.

(c) Conduct a test for weak instruments. Are your instruments weak? In what sense? Hint: See Section 4 in Andrews, Stock, and Sun (2019) .

We conduct the Olea and Pflueger (2013) test for weak instrument. The endogenous variable is `family_number_children` and the instrument is `second_birth_ismultiplebirth`, while we add three control variables, `first_child_age`, `first_child_is_girl`, `first_birth_ismultiplebirth`. The reported effective F-statistic was 5498.1354258. Since the authors suggested rule of thumb for a 5% test that the worst-case relative bias of 2SLS exceeds 10% is 23.1 for their corrected F-statistic, we **do reject** the null hypothesis that `second_birth_ismultiplebirth` is a weak instrument for `family_number_children`.

This test uses the first definition of Stock and Yogo of a weak instrument, that is, a instrument is said to be weak when the worst-case bias of two-stage least squares exceeds 10% of the worst-case bias of OLS.

(d) Report Anderson-Rubin confidence intervals. How do they compare to (b)?

We run the AR confidence interval for model 2, with control variables. The confidence interval found is [-0.4, -0.4], while this interval for the same model in item b) was [-0.4443, -0.3498], very close to each other.

(e) Compare your results with the estimates found in Question 1.

COMPARE TO ATE RESULTS FROM ITEM 1 C)

Part 2: Regression discontinuity design

For this part of the list, you have been provided with data from Amarante et al. (2016), who studies the effect of a cash-transfer program in Uruguay on health outcomes at birth. According to the authors, "the Uruguayan Plan de Atención Nacional a la Emergencia Social (PANES) was a temporary social assistance program targeted to the poorest 10 percent of households in the country, implemented between April 2005 and December 2007 ." Eligibility was defined via a baseline survey conducted with applicants. A probit model for the likelihood of falling below a critical per capita income level was estimated using baseline data, and households whose predicted probability exceeded some threshold were eligible to the program. However,

due to imperfect enforcement of the rules of the program, some noneligible mothers did actually receive the cash transfer, whereas some eligible mothers failed to do so.

You have been provided with a dataset where each entry corresponds to a pair (birth,mother) during the program duration. The treatment indicator variable is treat. The eligibility dummy is `eligible` $= \mathbf{1}\{\text{running} > 0\}$, where running is the predicted probability of falling to poverty, already subtracted of the threshold for program eligibility. File `dic_amarante.pdf` contains the description of additional variables in the dataset.

1. State the assumptions required for the identification of a treatment effect using the discontinuity described above. What do these assumptions mean in this context? What is the interpretation of the treatment effect identified under these assumptions?
2. Report a discontinuity plot between the running variable (x-axis) and program participation (y-axis), as well as a discontinuity plot between the running variable (x-axis) and low birthweight (variable `bajo2500`). What do these plots tell you?
3. Estimate the effect of program participation on low birthweight by local linear regression. Precisely state the bandwidth selection method used, the choice of kernel, as well as whether bias correction was employed. What is the first-stage relation, at the cutoff, between program eligibility and program participation? Is it statistically significant? Comment on your second stage results.
4. In order to assess the credibility of your empirical strategy, choose a variable which you may argue is predetermined and estimate the effect of program participation at the cutoff as in the previous item. What do you find?
5. Implement a manipulation test for the running variable in your setting. What do you find?
6. Do you think there are any potential threats to identification and/or estimation in your context? Can you think of any strategies to circumvent these?

References

Amarante, Verónica, Marco Manacorda, Edward Miguel, and Andrea Vigorito. 2016. "Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, Program, and Social Security Data." *American Economic Journal: Economic Policy* 8 (2): 1–43.

Andrews, Isaiah, James H Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11: 727–53.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Manski, Charles F. 2009. *Identification for Prediction and Decision*. Harvard University Press.

Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68 (4): 997–1010. http://www.jstor.org/stable/2999533.

Ponczek, Vladimir, and Andre Portela Souza. 2012. "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country." *Journal of Human Resources* 47 (1): 64–106.