Instrumental variables

Causal Inference, DPIR, University of Oxford

Tanisha Mohapatra & Fernando Sanchez Monforte

Week 4, Hilary Term 2025

Let's get started!

In this lab session, we will implement to concepts from this week's lecture[1].

Introduction

In this lab session[2] our focus will be on **instrumental variables** (IV) models - situations in which we don't have a random assignment of the treatment D_i, but we do have an as-if random variable Z_i, also known as an instrument. It is assumed to affect the treatment status D_i, but not the outcome Y_i, thanks to which we can still estimate the effects of our treatment of interest (as you already know from the lecture!). In the exercises below we will implement and validate instrumental regression models. In addition to coding, we will also do some brain-storming around the IV research design. In fact, doing independent research efficiently might be more about that than the coding per se.

We will use two examples of IV models In the first one, we will look at the literacy effect of encouraging children to watch the Sesame Street TV show. Here we will see one of two key uses of IV designs - accounting for non-compliance with binary treatment assignment. Specifically we will:

1. Explore the proportions of different compliance groups and their characteristics
2. Calculate ITT and LATE (Wald estimator)
3. Discuss the assumptions needed for the above calculations

In the second example, we will generalize IV approach by using 2SLS (two-stage least squares) models to estimate the electoral backlash against the wind turbines in Canada, instrumenting wind turbine construction with continuous measure of local wind speeds. We will:

1. Calculate first- and second-stage 2SLS equations
2. Estimate F-statistic
3. Implement 2SLS models controlling for covariates
4. Learn to use `ivreg` function from the `AER` package

Before starting this seminar

1. Create a folder called `lab4`.

[1] This session builds on materials kindly shared by Tom O'Grady, Jack Blumenau, and Julia de Romémont

[2] This practical session is building on materials kindly shared by Tom O'Grady and Jack Blumenau. The design is inspired by Marie-Lou Sohnius.

2. Download the data (`sesame_experiment.dta` and `Stokes.Rda`) from Canvas).
3. Save the data in our `lab4` folder.
4. Open the RMarkdown file (either download it from Canvas or start your own).
5. Set your working directory using the `setwd()` function or by clicking on "More". For example, this may look like this: `setwd(\"\~/Desktop/Causal Inference/2024/lab4\")`.
6. Prepare your R environment running the below.

Exercise 1. Sesame Street Experiment – Gladwell (2002) [3]

Can educational television programmes improve children's learning outcomes? Sesame Street is an American television programme aimed at young children. The creators of Sesame Street decided from the very beginning of the show's production that a central goal would to be educate as well as entertain its audience. As Malcolm Gladwell argued, "Sesame Street was built around a single, breakthrough insight: that if you can hold the attention of children, you can educate them". In addition to building the show around a carefully constructed educational curriculum, the show's producers also worked closely with educational researchers to determine whether the show's content was effectively improving its young viewers' numeracy and literacy skills.

The dataset contained in `sesame_experiment.dta` includes information on **240 children** who were randomly assigned to two groups. The treatment of interest here is watching Sesame Street, but clearly it is not possible to force children to watch a TV show or (perhaps even harder) to refrain from watching, and so **watching** the show cannot be randomized. Instead, in this study, researchers randomized whether children were **encouraged** to watch the show. More specifically, when the study was run in the 1970s, Sesame Street was on the air each day between 9am and 10am. The parents of children in the treatment group were encouraged to show Sesame Street to their children on a regular basis, while parents of the children in the control group were given no such encouragement. Because it is only encouragement that is randomized here, there is the possiblity of non-compliance – i.e. some children will not watch Sesame Street even though they are in the treatment condition, and some children will watch Sesame Street even though they are in the control condition. The list of variables in the dataset we are going to use is provided below.

Table 1: Variables of interest in the 'Sesame Street experiment' dataset from Gladwell

Variable	**Description**
`encour`	1 if the child was encouraged to watch Sesame Street, 0 otherwise
`watched`	1 if the child watched Sesame Street regularly, 0 otherwise
`letters`	the score of the child on a literacy test
`age`	age of the child (in months)
`female`	1 if the child is female, 0 otherwise

Load and explore the data first.

> Task 1.1. *(no coding)* In the context of this specific example, define the following unit types: 1. Compliers 2. Always-takers 3. Never-takers 4. Defiers

> Task 1.2. Calculate the proportion of children in the treatment group who did not watch Sesame Street. Calculate the proportion of children in the control group who did watch Sesame Street. What type of non-compliance occured in this experiment? Are the different compliance subgroups comparable with respect to age and gender?

[3] Gladwell M. (2002). [The Tipping Point: How Little Things Can Make a Big Difference]https://www.amazon.co.uk/Tipping-Point-Little-Things-Difference/dp/0349113467)." *Abacus*

Hint: You might find the `table()` and `prop.table()` functions helpful here.

Task 1.3 Calculate the proportion of compliers in this experiment. Which assumptions are required for us to identify this quantity?

Hint: We can calculate the proportion of compliers via $E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = \bar{D}_{Z_i=1} - \bar{D}_{Z_i=0}$

Task 1.4. Calculate the Intention-to-Treat effect (ITT). Is it substantive? What about statistical significance? Ultimately, how would you interpret ITT here?

Hint: We can calculate the ITT via $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = \bar{Y}_{Z_i=1} - \bar{Y}_{Z_i=0}$

Task 1.5. Now calculate the local average treatment effect (LATE). What does the LATE estimate? Estimate the LATE for this example.:

Hint: Use Wald estimator, manual 2SLS using `lm`, or `ivreg` from the `AER` package (the latter two will be discussed in greater detail in Exercise 2)

Task 1.6. *(no coding)* In addition to 1) monoticity and 2) independence of the instrument assumptions, there are 2 other assumptions that we need to make to claim the estimates are causal. What are those? Do you think they are satisfied?

Task 1.7. *(no coding)* You have now estimated two treatment effects: the ITT and LATE. Which is of greater interest to the TV show's producers?

Exercise 2. Electoral Backlash against Climate Policy – Stokes (2016) [4]

Leah Stokes (2016) studies whether governments are punished at the ballot box for building wind farms. Construction of such projects is a policy that can help to mitigate climate change but it might also impose costs on the communities where turbines are sited. She looks at Ontario in Canada, where from 2009 the centre-left provincial government removed local communities' right to make planning decisions on the building of wind turbines. Instead, decision-making was centralised and turbines were imposed by the government. It built turbines in places where they would generate the most electricity: in places with higher prevailing wind speeds. In general certain broad areas are better-suited for turbines (rural and elevated places, and areas closer to the windy Great Lakes), Stokes argues that within these areas wind speed varies at random at the local level. Local areas with high wind speeds should not be more supportive of the government than local areas with low wind speeds. This is therefore a natural experiment where wind speed is an instrument that randomly encouraged the government to site turbines in particular places. Her outcome of interest is change in support for the incumbent government from 2007 (before the wind farm policy) to 2011 (after it began) at a highly localised level known as precincts in Canada, which typically contain around 300 voters. Using GIS software, she geo-located all wind turbines that were built or proposed in the period and matched them to precincts, where she collected voting data, localised prevailing wind speeds, and background covariates. The dataset for this question is contained in `Stokes.Rda` and contains the following variables.

[4] Stokes, L. (2016). Electoral Backlash against Climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy." *American Journal of Political Science*, 60 (4), pp. 958-974

Table 2: Variables of interest from Stokes 2016

Variable	Description
`chng_lb`	**outcome** - percentage point change in support for the incumbent government, 2007-2011
`prop_3km`	**treatment** - dummy indicating whether a wind turbine was built or proposed within 3km (1), or not (0)
`avg_pwr_log`	**instrument** - prevailing wind speed in the precinct, logged
`longitude`	geographical longitude (East-West) of the precinct
`latitude`	geographical latitude (South-North) of the precinct
`ed_id`	the broader district within which the precinct is located
`mindistlake`	distance to the Great Lakes in km
`mindistlake_sq`	distance to the Great Lakes in km, squared

Task 2.1. Assess whether wind speed can be considered to be as-if randomly assigned geographically, by regressing `avg_pwr_log` on all of the geographical covariates. What do you conclude?

Hint: Remember to use factor() for the `ed_id` variable

Task 2.2. Estimate the first-stage relationship between `prop_3km` and `avg_pwr_log` using a regression with no added covariates. Interpret the results.

Task 2.3. *(no coding)* Ok, is that all we need for the first stage? What about all the other covariates in the dataset? In fact, Stokes (2016) does use geographic controls for both first and second stage models. Why do you think that is? Even if we do that, do you think we control for every relevant variable?

Task 2.4. Given the discussion above, re-estimate the first-stage relationship between `prop_3km` and `avg_pwr_log`, this time with a full set of geographic controls. Fully interpret the results. Does it differ from your results in Task 2.2 above?

Task 2.5. Conduct an F-test for the strength of the `avg_pwr_log` instrument.

Hint: use the function `waldtest` in the `lmtest` library. Your code should take the form `waldtest(model1,model2)` , where model1 and model2 are the names of estimated regression models with and without the instrument

Given the above, we might also re-calculate F-statistics manually. How?

Hint: Applying `anova` function on your model will give you several summary statistics, including the residual sum of squares, which you can then save as separate value "'

Task 2.6. Knowing the above, try estimating the Local Average Treatment Effect using two-stage least squares specifying both the first and second stage equations (with covariates). How would interpret the results?

Hint: Extract fitted values from the first stage using `fitted.values()` and use them as explanatory variable in the second stage regression.

Task 2.7. Now estimate the Local Average Treatment Effect of `prop_3km` on `chng_lib` using two-stage least squares with `avg_pwr_log` as the instrument and the full set of geographic controls. Interpret the coeffcient on `prop_3km` and its statistical significance precisely.

Hint: Use `ivreg` in the AER library. Your code should take the form: `ivreg(outcome ~ treatment + covariates | instrument + covariates)`

Task 2.8. *(no coding)* What about other assumptions behind instrumental variables - the independence of the instrument and exclusion restriction?