Introduction to R and econometrics - Part II

Ryan Hawthorne
Acacia Economics
CCRED (UJ)

23 July, 2021

Causal effects

Some methods to consistently estimate causal effects

We will discuss several methods that could be used to overcome endgeniety problems in order consistently estimate regression parameters that describe causal effects (like the slope of a demand function).

1. Conduct a randomized experiment.
2. Add control variables
3. Use instrumental variable estimation.

Conduct a Randomized Experiment

- The ideal method to estimate a causal effect is to run a randomized experiment. We have already illustrated this.
- Randomised experiments are often called the *Scientific Gold Standard* to establish causal effects. They are for example required by regulators when a pharmaceutical company wants to establish that a new drug has positive effects on patients.
- However, it is not always possible, or too costly, to run a randomized experiment. We thus learn the other approaches

Hypothesis tests: Null hypothesis

- A hypothesis test consists of a **null hypothesis** H_0 and a corresponding **alternative hypothesis** H_1 about some features of a data generating process. Examples for hypotheses for a linear regression model:
    - H_0: $\beta_1 = 0$, $H_1 : \beta_1 \neq 0$
    - H_0: The explanatory variable x_k is exogenous, $H_1 : x_k$ is endogenous
    - H_0: The disturbance ε is not auto-correlated, H_1: ε is auto-correlated

Example: t-test for a regression coefficient

- Consider a linear regression model $y = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K + \varepsilon$ that satisfies a multiple regression equivalent to assumptions (A1)-(A4) and the null hypothesis:

$$H_0 : \beta_k = 0$$

- Every hypothesis test is based on a **test statistic** that can be computed from the data. In our example, it has *t-value*:

$$t_k = \frac{\hat{\beta}_k}{\hat{sd}(\hat{\beta}_k)}$$

- We can also view a test statistic as a random variable. Here t_k is a transformation of the random variable ε and the explanatory variables.

- Key of every hypothesis test is that one knows the distribution

P-values and significance levels

- The p-value measures the probability to find the realized or more extreme test statistic if H0 is true (see plot above).
- One often considers critical levels of the p-value like 5% or 1%, which are called significance levels.
- We say we can reject the H0 at significance level α if the p-value is smaller than α,
  - e.g. if we have p-value=0.043 we can reject H0 at a significance level of 5%.
- Significance levels are often marked with one or several stars ** in regression outputs.