University of Essex
2015-16

Department of Government
TA: Stavros Poupakis

GV903: Advanced Research Methods

Class 14 Instrumental Variables

Draw a random sample of 1,000 observations from $x_{i1}^* \sim N(0, 2^2)$. Generate $\epsilon_i = 2x_{i1}^* + N(0, 3^2)$.

Simulate also $z_i \sim N(0, 5^2)$ and create $x_{i1} = x_{i1}^* + z_i$. Simulate also $x_{i2} \sim \chi^2$ with 2 d.f.

Generate then y_i as

$$y_i = 5 + 2x_{i1} + 3x_{i2} + \epsilon_i$$

1. Check the mean and variance of all the variables you generated.
2. What do you expect to be the value of the covariance between x_1^* and ϵ? Check with R.
3. What do you expect to be the value of the covariance between z and ϵ? Check with R.
4. What do you expect to be the value of the covariance between x_1 and z? Check with R.
5. Calculate the correlation of $x1$ and ϵ in R *without* using the command `cor`. Do the same for x_2 and ϵ, and z and ϵ.
6. Fit an OLS of y on x_1 and x_2.
7. Fit a Two Stage Least Squares using two OLS steps.
8. Fit the 2SLS using the `ivreg` function from the *AER* package. Make sure you use the options `vcov = sandwich, diagnostics = TRUE`.
9. Perform a Monte Carlo experiment to assess the finite sample performance of OLS and 2SLS. Use a loop for that. Look at the following example and adjust it to your code:

```
M <- 1000
stats <- matrix(NA, nrow=M, ncol=2)
for (i in 1:M){
  n <- 200
  x <- rnorm(n)
  stats[i,1] <- mean(x)
  stats[i,2] <- sd(x)
}
mean(stats[,1])
mean(stats[,2])
apply(stats, 2, mean)
apply(stats, 2, sd)
```

10. Violate the assumption $Cov(z, \epsilon) = 0$ for the validity of the instrument; see what happens.
11. Violate the assumption for $Cov(x_1, z)$ to make the instrument weak; see what happens.
12. Do we care about $Cov(x_2, z)$? In other words, does it matter if the exogenous variable is correlated with the instrument?