Econometrics - Advanced Methods

Linear regression model (MC-Simulation and IV)

A. Use the do-file "EAM2019 - class 2A - MCsim.do" to perform a Monte Carlo simulation for OLS.

1. Fill the gaps in the do-file.

2. Start with the easiest case where x and u are normally distributed. Vary the sample size and/or the variance of x and u, and compare the observed variance of $\widehat{\beta}$ with the asymptotic variance under homoskedasticity. Also, have a look at the empirical distribution of $\widehat{\beta}$ (`kdensity bx, normal`).

3. Now, try different distributions for x and u. For example, the uniform distribution (`runiform(n,1)`) or the Cauchy distribution. What do you observe?

B. Use the dataset "iv.dta" and suppose we are only interested in the effect of x_1. Consider the following population model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \,.$$

z is a potential instrumental variable for x_1. The following conditions hold jointly: $cov(x_1, u) \neq 0$, $cov(x_2, u) = 0$ and $cov(z, u) = 0$.

1. Is z a relevant instrument for x_1? What is the "usual" critical value of the F-statistic when it comes to weak instruments? As we have only one instrument in this application, the critical value of the F-statistic can be linked to a critical value for the t-test. Derive this t-test critical value and its corresponding p-value (use the large-sample approximation).

2. Use OLS and 2SLS to estimate β_1. Which one gives you a consistent estimate? And why?

3. How can you test the exogeneity of x_1?

4. Now, assume we don't observe x_2. Under what additional assumption is the IV estimate of β_1 consistent?

C. Consider the following model

$$y = X\beta + u$$
$$X = Z\pi + v\,.$$

Use the do-file "EAM2019 - class 2C - Mata 2SLS.do" to estimate 2SLS in Mata.

1. To make OLS inconsistent, we need correlation b/w u and v (a part of X). Wlog, let u, v, w be independently standard normal distributed. Define

$$u = \rho v + \sqrt{1 - \rho^2}\, w\,.$$

   Proof that $Corr(u, v) = \rho$.

2. Fill the gaps in the do-file. Use again the projection matrix, although it's not an efficient way to estimate 2SLS.

3. Increase the sample size to 15,000 and compare the calculation speed of your Mata file with Stata's `ivregress 2sls`.

4. Can you think about a more efficient way to estimate 2SLS (in particular, without defining a $n \times n$ matrix)?