Econ 797B Problem Set 3: Instrumental Variables Estimation

Due Friday November 20, 2020

Consider the following simulated dataset with a single outcome, a single regressor and multiple instruments

$y_i = \beta x_i + \eta_i$
$x_i = \sum_{j=1}^{Q} \pi_j z_{ij} + \xi_i$
with
$\beta = 1, \pi_1 = 0.1, \pi_j = 0 \forall j > 1,$
$\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} | Z \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \right)$

Problem 1: Weak First Stage Simulations - OLS, 2SLS, CLR, LIML, [AND LASSO: optional]

Simulate this 1,000 times each for Q=1, 10, 20:

- Draw samples of 1000 from the DGP above, and estimate the following:
  - a) OLS
  - b) 2SLS (conventional confidence interval)
  - c) 2SLS (CLR confidence interval using `weakiv` command (need to SSC install))
  - d) LIML
  - (e) use (post) LASSO for instrument selection in first stage and reduced form like in Belloni et al. (2012), or using Stata 16's `poivregress` command. [Optional]
- In each of these cases, compute: (i) the bias ($\hat{\beta} - 1$), (ii) whether you reject the (correct) null that $\beta = 1$, and store in a matrix. Make the matrix into a dataset and report (i) the median bias (ii) Type 1 error (probability of incorrectly rejecting the null $\beta = 1$). For post-LASSO, keep track of what share of simulations don't pick any instrument, and then report the bias and Type 1 error for the simulations that do pick some instruments. Report these in a clearly formatted table for different Q and different estimators. Discuss your findings.