Statistics 215B Assignment 4 Due: 5pm Mon 19 Feb 2024

For this assignment you will use simulation to study the performance of OLS and IVLS coefficient estimators, as well as two different estimators of the error variance. Consider the model

$$\begin{aligned} Y_i &= X_i \beta + \epsilon_i, \\ X_i &= U_i + 2V_i + \delta_i. \end{aligned} \tag{1}$$

Everything in sight is scalar. The vectors

$$(U_i, V_i, \epsilon_i, \delta_i),\ i = 1, \ldots, n,$$

are iid across $i$. Each vector is normal with mean zero. For each $i$ we take the three random objects (1) $U_i$, (2) $V_i$, and (3) $(\epsilon_i, \delta_i)$ to be mutually independent; furthermore,

$$\begin{aligned} \mathrm{Var}(U_i) &= \mathrm{Var}(V_i) = 1, \\ \mathrm{Var}(\epsilon_i) &= \mathrm{Var}(\delta_i) = \sigma^2, \text{ and} \\ \mathrm{Cov}(\epsilon_i, \delta_i) &= \rho. \end{aligned}$$

In (1), the variable $X_i$ is endogenous (explain why), and $(U_i, V_i)$ are instruments (explain why).

Suppose $\beta = 3$, $\sigma^2 = 1$, and $\rho = 3/4$. For each of 1,000 simulation runs, generate $n = 100$ independent realizations of the vector $(U_i, V_i, \epsilon_i, \delta_i, X_i, Y_i)$ according to (1) and (2). Use OLS to obtain the estimate $\hat{\beta}_{\text{OLS}}$, and IVLS to obtain $\hat{\beta}_{\text{IVLS}}$. Plot the histogram for each estimator; report the mean, SD, and RMSE in each case. What are the relative merits of OLS versus IVLS? For each simulation, estimate the error variance $\sigma^2$ in two ways: first using the residuals obtained from plugging $\hat{\beta}_{\text{IVLS}}$ into (1), then using the residuals from the transformed equation

$$(Z'Z)^{-1/2} Z'Y = (Z'Z)^{-1/2} Z'X\beta + \eta\ .$$

Here $Z$ is the $n \times 2$ matrix of instruments, $X$ is the $n \times 1$ design matrix, and $Y$ is the $n \times 1$ vector of responses. What is the appropriate denominator in each case? Plot the histograms for the two estimators and report sample means and SDs. Comment briefly.