PS #8 (Due 20 March, 2019)

1. Consider the following regression model with a measurement error problem in PS#7:

$$y_i = 1 + x_i^* + z_i + u_i$$

where

$$\begin{pmatrix} x_i^* \\ z_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$$

and $x_i = x_i^* + v_i$ where $v_i \sim N(0,1)$ where v_i and (x_i, z_i, u_i) are independent. In this question, in addition to x_i, assume that there is a second measurement of x_i^*, i.e.

$$x_{2i} = x_i^* + v_{2i},$$

where v_{2i} and (x_i, z_i, u_i, v_i) are independent.

(a) Show that $(1, x_{2i}, z_i)$ constitute valid instrumental variables for the linear regression model

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i - \beta v_i$$

where $\alpha = 1$, $\beta = 1$, $\gamma = 1$, by showing

$$E\begin{pmatrix} u_i \\ x_{2i}u_i \\ z_i u_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad E\begin{pmatrix} 1 & x_i & z_i \\ x_{2i} & x_{2i}x_i & x_{2i}z_i \\ z_i & z_i x_i & z_i^2 \end{pmatrix} \text{ is invertible.}$$

(b) Explain why the above two conditions are important for the validity of $(1, x_{2i}, z_i)$ as instrumental variables.

(c) Pretending that you did not know the true coefficients, estimate them by IV method.

(d) Write R and Python code to demonstrate by simulation that the IV estimator is consistent and asymptotically normal.

(e) Again, pretending that you did not know the true coefficients, estimate them by the two stage least squares (2SLS) method not by using the 2SLS command, but literally running the least squares in two stages; first regress each of the regressors on instrumental variables and obtaining the predicted values of the regressors by instrumental variables, then in the second stage regressing the dependent variables on the predicted values. You should obtain exactly the same estimate with the IV estimator.

(f) In general the IV estimator and the 2SLS estimator are numerically the same when the number of the instrumental variables and the number of regressors in the model are the same. To see this, show the following:

i. Recall that

$$\hat{\beta}_{2SLS} = (\mathbf{X}'P_Z\mathbf{X})^{-1}\mathbf{X}'P_Z\mathbf{Y} \quad \text{and} \quad \hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y}$$

Write the 2SLS estimator using matrices $\mathbf{X}$, $\mathbf{Z}$, and $\mathbf{Y}$.

ii. Show that the sizes of matrices $\mathbf{X}'\mathbf{Z}$ and $\mathbf{Z}'\mathbf{X}$ are both $K \times K$ and when the IV is well defined they are invertible.

iii. Note that the above does not hold when there are more than K instrumental variables.

iv. Use the result (ii) to show that $\hat{\beta}_{2SLS} = \hat{\beta}_{IV}$.