AE HW 5

5.1. In this problem you are to establish the algebraic equivalence between 2SLS and OLS estimation of an equation containing an additional regressor. Although the result is completely general, for simplicity consider a model with a single (suspected) endogenous variable:

$$y_1 \overset{x\beta}{=} \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + u_1$$

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2$$

For notational clarity, we use y_2 as the suspected endogenous variable and $\mathbf{z}$ as the vector of all exogenous variables. The second equation is the reduced form for y_2. Assume that $\mathbf{z}$ has at least one more element than $\mathbf{z}_1$.

We know that one estimator of $(\boldsymbol{\delta}_1, \alpha_1)$ is the 2SLS estimator using instruments $\mathbf{x}$. Consider an alternative estimator of $(\boldsymbol{\delta}_1, \alpha_1)$: (a) estimate the reduced form by OLS, and save the residuals $\hat{v}_2$; (b) estimate the following equation by OLS:

$$y_1 = \mathbf{z}_1\boldsymbol{\delta}_1 + \alpha_1 y_2 + \rho_1\hat{v}_2 + error \tag{5.52}$$

Show that the OLS estimates of $\boldsymbol{\delta}_1$ and α_1 from this regression are identical to the 2SLS estimators. [Hint: Use the partitioned regression algebra of OLS. In particular, if $\hat{y} = \mathbf{x}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{x}_2\hat{\boldsymbol{\beta}}_2$ is an OLS regression, $\hat{\boldsymbol{\beta}}_1$ can be obtained by first regressing $\mathbf{x}_1$ on $\mathbf{x}_2$, getting the residuals, say $\ddot{\mathbf{x}}_1$, and then regressing y on $\ddot{\mathbf{x}}_1$; see, for example, Davidson and MacKinnon (1993, Section 1.4). You must also use the fact that $\mathbf{z}_1$ and $\hat{v}_2$ are orthogonal in the sample.]

Following given steps

$$y_1 = \underbrace{z_1\delta_1 + \alpha_1 y_2}_{\underset{x_1}{[z_1, y_1]}\ \underset{\beta_1}{\begin{bmatrix}\delta_1\\ \alpha_1\end{bmatrix}}} + \underset{\beta_2}{\rho_1}\ \underset{x_2}{\hat{v}_2} + error$$

Step 1. $L(z_1, y_2 \mid \hat{v}_2)$

Since $\hat{v}_2 \perp z_1$

$$L[z_1 \mid \hat{v}_2] = 0 \cdot \hat{v}_2 + z_1$$

Since $y_2 = \hat{y}_2 + \hat{v}_2$

$$\therefore \begin{bmatrix} z_1 \\ y_2 \end{bmatrix} = [\hat{v}_2\ \hat{v}_2]\begin{bmatrix} 0 \\ 1 \end{bmatrix} + \underset{\ddot{x}_1}{\underset{|||}{\begin{bmatrix} z_1 \\ \hat{y}_2 \end{bmatrix}}}$$

Step 2 $L(y_1 \mid \ddot{x}_1)$

$$= L(y_1 \mid z_1, \hat{y}_2)$$

2SLS Steps

when we treat y_2 as the endogeneous variable in the system then we take the following two steps:

Step 1. $L(y_2 \mid z)$

$\Rightarrow \hat{y}_2$

Step 2. $L(y_1 \mid z_1, \hat{y}_2)$

$=$

5.9. Suppose that the following wage equation is for working high school graduates:

$$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 twoyr + \beta_4 fouryr + u$$

where *twoyr* is years of junior college attended and *fouryr* is years completed at a four-year college. You have distances from each person's home at the time of high school graduation to the nearest two-year and four-year colleges as instruments for *twoyr* and *fouryr*. Show how to rewrite this equation to test H_0: $\beta_3 = \beta_4$ against H_0: $\beta_4 > \beta_3$, and explain how to estimate the equation. See Kane and Rouse (1995) and Rouse (1995), who implement a very similar procedure.

$\theta_4 \overset{def}{=} \beta_4 - \beta_3 \Rightarrow \beta_4 = \beta_3 + \theta_4.$

Plug in the specification

$$\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 (twoyr + fouryr) + \theta_4 fouryr + u$$
$$= \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 totcoll + \theta_4 fouryr + u$$

$totcoll = twoyr + fouryr$

Now. just estimate the latter equation by 2SLS.

w/. exper, exper², dist2yr. dist4yr as IV's