Problem Set #3

Danny Edgel
Econ 710: Economic Statistics and Econometrics II
Spring 2021

February 13, 2021

Collaborated with Sarah Bass, Emily Case, Katherine Kwok, Michael Nattinger, and Alex Von Hafften

Question 1

(i) Before showing this equivalence, let us first observe $\mathbb{E}[ZX']$ and $\mathbb{E}[ZZ']$:

$$\mathbb{E}[ZX'] = \begin{pmatrix} \mathbb{E}[Z_1 X_1] & \mathbb{E}[X_2' Z_1] \\ \mathbb{E}[X_2 X_1] & \mathbb{E}[X_2 X_2'] \end{pmatrix}, \mathbb{E}[ZZ'] = \begin{pmatrix} \mathbb{E}\left[Z_1^2\right] & \mathbb{E}[X_2' Z_1] \\ \mathbb{E}[X_2 Z_1] & \mathbb{E}[X_2 X_2'] \end{pmatrix}$$

Note that the only difference between these two matrices is the first column. Further, since X_2 is assumed non-zero, $\mathbb{E}[X_2 X_2']$ is invertible. Thus, we know that the rank of the last $k-1$ columns of $\mathbb{E}[ZX']$ and $\mathbb{E}[ZZ']$ is $k-1$ so long as $\mathbb{E}[X_2' Z_1] \neq 0$. We will show each side of the "if and only if" condition separately.

(a) If $\mathbb{E}[ZZ']$ is invertible and $\pi \neq 0$, then we know that $\mathbb{E}[ZZ']^{-1} y = 0$ iff $y = 0$. Since $\pi_1 \neq 0$, $\mathbb{E}[ZX_1] \neq 0$. Thus, the first column of $\mathbb{E}[ZX'] \neq 0$. Therefore, $\mathbb{E}[ZX']$ is invertible.

(b) If $\mathbb{E}[ZX']$ is invertible, then rank($\mathbb{E}[ZX']$)=k, so $\pi_1 \neq 0$. This implies that $\mathbb{E}\left[Z_1^2\right] \neq 0$, so $\mathbb{E}[ZZ']$ is invertible.

$\therefore$ $\mathbb{E}[ZX']$ is invertible $\iff$ $\mathbb{E}[ZZ']$ is invertible and $\pi \neq 0$ ■

(ii) Given our asymptotic distribution of $\hat{\beta}^{IV}$, if we assume homoskedasticity, we have:

$$\Omega = \sigma_U^2 \mathbb{E}[ZX']^{-1} \mathbb{E}[ZZ'] \mathbb{E}[XZ']^{-1}$$

Using the block inversion formula, we can solve:

$$\mathbb{E}\left[Z'X\right]^{-1} = \frac{1}{k}\begin{pmatrix} 1 & -\mathbb{E}\left[Z_1 X_2'\right]\mathbb{E}\left[X_2 X_2'\right]^{-1} \\ . & . \end{pmatrix}$$

$$\mathbb{E}\left[XZ'\right]^{-1} = \frac{1}{k}\begin{pmatrix} 1 & . \\ -\mathbb{E}\left[Z_1 X_2\right]\mathbb{E}\left[X_2 X_2'\right]^{-1} & . \end{pmatrix}$$

$$\text{Where } k = \mathbb{E}\left[Z_1 X_1\right] - \mathbb{E}\left[X_2' X_1\right]\mathbb{E}\left[X_2 X_2'\right]^{-1}\mathbb{E}\left[X_2 Z_1\right] = \mathbb{E}\left[X_1 \widetilde{Z}_1\right]$$

And where neither the second row of $\mathbb{E}\left[Z'X\right]^{-1}$ nor the second column of $\mathbb{E}\left[XZ'\right]^{-1}$ enter Ω_{11}. Then,

$$\begin{aligned}\Omega_{11} &= \frac{\sigma_U^2}{\mathbb{E}\left[X_1 \widetilde{Z}_1\right]^2}\left(\mathbb{E}\left[Z_1^2\right] - \mathbb{E}\left[Z_1 X_2'\right]\mathbb{E}\left[X_2 X_2'\right]^{-1}\mathbb{E}\left[X_2 Z_1\right]\right) \\ &= \frac{\sigma_U^2}{\mathbb{E}\left[X_1 \widetilde{Z}_1\right]^2}\mathbb{E}\left[\widetilde{Z}_1\right]^2 = \frac{\sigma_U^2 \widetilde{Z}_1}{\mathbb{E}\left[\widetilde{Z}_1^2\right]\pi_1^2}\end{aligned}$$

(iii) $(\pi_1, \pi_2)'$ is the vector of coefficients from a regression of X_1 on Z, and $\widetilde{Z}_1$ is the residualized value of Z_1 from said regression.

(iv) First, recognize that Ω_{11} can be rewritten as:

$$\frac{\sigma_U^2 \mathbb{E}\left[\widetilde{Z}_1^2\right]}{\mathbb{E}\left[\widetilde{Z} Z_*\right]^2}$$

Then, by deploying Cauchy-Schwarz, we can derive:

$$\begin{aligned}\Omega_{11} = \frac{\sigma_U^2 \mathbb{E}\left[\widetilde{Z}_1^2\right]}{\mathbb{E}\left[\widetilde{Z} Z_*\right]^2} &\geq \frac{\sigma_U^2 \mathbb{E}\left[\widetilde{Z}_1^2\right]}{\mathbb{E}\left[\widetilde{Z}\right]^2 \mathbb{E}\left[Z_*\right]^2} \\ &= \frac{\sigma_U^2}{\mathbb{E}\left[\mathbb{E}\left[Z_*\right]^2\right]}\end{aligned}$$

We could acheive $\Omega_{11} = \frac{\sigma_U^2}{\mathbb{E}\left[\mathbb{E}\left[Z_*\right]^2\right]}$ if $\widetilde{Z} = \mathbb{E}\left[Z_*\right]$.

(v) Let $X_2 = 1$. Then π_1 is simply the coefficient from a regression of X_1 on

Z_1 that includes a constant term, so:

$$\widetilde{Z}_1 = Z_1 - \mathbb{E}[Z_1] \Rightarrow \mathbb{E}\left[\widetilde{Z}_1\right] = Var(Z_1)$$

$$\mathbb{E}\left[X_1\widetilde{Z}_1\right] = Cov(X_1, \widetilde{Z}_1)$$

$$\Omega_{11} = \frac{\sigma_U^2}{\mathbb{E}\left[X_1\widetilde{Z}_1\right]^2}\mathbb{E}\left[\widetilde{Z}_1\right]^2 \Rightarrow \Omega_{11}$$

$$\therefore \frac{\Omega_{11}}{\sigma_U^2} = \frac{Var(Z_1)}{Cov(X_1, \widetilde{Z}_1)^2}$$

Question 2

(i) Let us begin by simplifying $\mathbb{E}[h(Z)(Y - X\beta)]$:

$$\begin{aligned}\mathbb{E}[h(Z)(Y - X\beta)] &= \mathbb{E}[h(Z)(X\beta_1 + U - X\beta)] \\ &= \mathbb{E}[h(Z)\mathbb{E}[U|Z]] + \mathbb{E}[h(Z)X(\beta_1 - \beta)] \\ &= (\beta_1 - \beta)\mathbb{E}[h(Z)X]\end{aligned}$$

If $\mathbb{E}[h(Z)X] \neq 0$, then $(\beta_1 - \beta)\mathbb{E}[h(Z)X] = 0$ if and only if $\beta = \beta_1$.

(ii) Suppose $\mathbb{E}[h(Z)X] \neq 0$. Then, we can derive $\hat{\beta}_1^h$ using the equality from (i):

$$\mathbb{E}[h(Z)(Y - X\beta)] = 0 \Rightarrow \beta\mathbb{E}[h(Z)X] = \mathbb{E}[h(Z)Y]$$

$$\beta = \frac{\mathbb{E}[h(Z)Y]}{\mathbb{E}[h(Z)X]} \Rightarrow \hat{\beta}_1^h = \frac{\sum_{i=1}^n h(Z_i)Y_i}{\sum_{i=1}^n h(Z_i)X_i}$$

(iii) By the central limit theorem,

$$\hat{\beta}_1^h = \frac{\sum_{i=1}^n h(Z_i)Y_i}{\sum_{i=1}^n h(Z_i)X_i} = \frac{\sum_{i=1}^n h(Z_i)(X_i\beta_1 + U)}{\sum_{i=1}^n h(Z_i)X_i} = \beta_1 + \frac{\sum_{i=1}^n h(Z_i)U}{\sum_{i=1}^n h(Z_i)X_i}$$

$$\sqrt{n}\left(\hat{\beta}_1^h - \beta_1\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n h(Z_i)U}{\frac{1}{n}\sum_{i=1}^n h(Z_i)X_i} \rightarrow_d \frac{\mathbb{E}[h(Z)U]}{\mathbb{E}[h(Z)X]}\mathcal{N}(0,1) = \mathcal{N}\left(0, \Omega^h\right)$$

$$\Omega^h = \frac{\mathbb{E}\left[h(Z)^2U^2\right]}{\mathbb{E}[h(Z)X]^2}$$

(iv) Using the Cauchy-Schwarz inequality, we can show:

$$\begin{aligned}
\Omega^h &= \frac{\mathbb{E}\left[h(Z)^2U^2\right]}{\mathbb{E}\left[h(Z)X\right]^2} = \frac{\mathbb{E}\left[h(Z)^2\mathbb{E}\left[U^2|Z\right]\right]}{\mathbb{E}\left[h(Z)\mathbb{E}\left[X|Z\right]\right]^2} \\
&\geq \mathbb{E}\left[\frac{h(Z)^2\mathbb{E}\left[U^2|Z\right]}{h(Z)^2\mathbb{E}\left[X|Z\right]^2}\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[U^2|Z\right]}{\mathbb{E}\left[X|Z\right]^2}\right] \\
\Rightarrow \Omega^h &\geq \mathbb{E}\left[\frac{\mathbb{E}\left[X|Z\right]^2}{\mathbb{E}\left[U^2|Z\right]}\right]^{-1}
\end{aligned}$$

This lower bound is achieved by $h(Z) = \frac{\mathbb{E}[X|Z]}{\mathbb{E}[U^2|Z]}$:

$$\begin{aligned}
\Omega^h &= \frac{\mathbb{E}\left[\left(\frac{\mathbb{E}[X|Z]}{\mathbb{E}[U^2|Z]}\right)^2U^2\right]}{\mathbb{E}\left[\frac{\mathbb{E}[X|Z]}{\mathbb{E}[U^2|Z]}X\right]^2} = \frac{\mathbb{E}\left[\frac{\mathbb{E}[X|Z]^2}{\mathbb{E}[U^2|Z]^2}\mathbb{E}\left[U^2|Z\right]\right]}{\mathbb{E}\left[\frac{\mathbb{E}[X|Z]}{\mathbb{E}[U^2|Z]}\mathbb{E}\left[X|Z\right]\right]^2} \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[X|Z\right]^2\right]}{\mathbb{E}\left[\frac{\mathbb{E}[X|Z]^2}{\mathbb{E}[U^2|Z]}\right]^2} = \frac{1}{\mathbb{E}\left[\frac{\mathbb{E}[X|Z]^2}{\mathbb{E}[U^2|Z]}\right]} \\
&= \mathbb{E}\left[\frac{\mathbb{E}\left[X|Z\right]^2}{\mathbb{E}\left[U^2|Z\right]}\right]^{-1}
\end{aligned}$$

Question 3

The attached Matlab file, `edgel_ps3.m`, calculates $\hat{\beta}_1^{2SLS}$ and $\hat{V}_\beta$ using formula (12.40) of Professor Hansen's textbook. The results for β_1 are displayed below, rounded to the nearest thousandth.

$$\begin{aligned}
\beta_1^{2SLS} &= 0.108 \\
\text{SE} &= 0.020
\end{aligned}$$

The code used is provided at the end of this file.

```matlab
%{
    This file is used to perform the tasks required for question 3 of
    problem set 3 for the first quarter of ECON 710

    Date created:  13 Feb 2021
    Last modified: 13 Feb 2021
    Author: Danny Edgel
%}

% clean workspace
clc; clear

%% Step 1: Import and prepare Angrist & Krueger (1991) data

% read data from .csv file
dat = readtable('AK91.csv');

% Extract X1 and vectors
Y  = dat.lwage;
X1 = dat.educ;

% generate X2 matrix -- for state and year of birth dummies, remove states
% and years with no observations (otherwise X won't be invertible)
v = {'sob', 'yob'};

for i = 1:length(v)
    x.(v{i})    = dummyvar(dat.(v{i}));
    count       = sum(x.(v{i}));
    x.(v{i})    = x.(v{i})(:, count>0);
end

% (exclude first dummy of each dummy matrix)
X2 = [ ones(size(X1)), x.yob(:,1:(end-1)), x.sob(:,1:(end-1)) ];

% save X separately, for coding ease in step 2
X = [ X1, X2 ];

% generate Z matrix from qob and X2 (exclude first quarter)
qob_dummies = dummyvar(dat.qob);

```

```
Z = [ qob_dummies(:,2:4), X2 ];

% save sample size
n = size(X1, 1);

%% Step 2: calculate 2SLS estimator for beta_1 and its heteroskedasticity-
%   robust standard error from (12.40) in Hansen's textbook

% estimate beta
bhat =  (X'*Z/(Z'*Z)*Z'*X)\(X'*Z/(Z'*Z)*Z'*Y);

% estimate the constituent matrices of the standard error
Q_zz = (1/n)*(Z'*Z);
Q_xz = (1/n)*(X'*Z);
ehat = Y - X*bhat;
Ohat = 0*Q_zz;
for i=1:n; Ohat = Ohat + (1/n)*Z(i,:)'*Z(i,:)*ehat(i)^2; end

% calculate standard error
Vb = ((Q_xz/Q_zz*Q_xz')\(Q_xz/Q_zz*Ohat/Q_zz*Q_xz')/(Q_xz/Q_zz*Q_xz'))/n;


%% Step 3: output results

% initialize LaTeX file
filename = 'q3.tex';

if exist(filename, 'file')==2
  delete(filename);
end
file1 = fopen(filename,'w');

% output bhat_1 and its heteroskedasticity-robust SE
tex = ...
    ['\\begin{align*}\n',...
    '\\beta^{2SLS}_1 &= %4.3f \\\\ \n',...
    '\\text{SE}      &= %4.3f  \n', ...
    '\\end{align*}'];
fprintf(file1,tex,...
    round(bhat(1), 3), round(sqrt(Vb(1, 1)), 3));
fclose(file1);
```