SP25 8732: Homework 2

Danbo CHEN

February 3, 2025

1 Question 1

(a) Solution:

```
. /// 1.A
> reg price sqrft bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(2, 85)        =     72.96
       Model |  580009.152         2  290004.576   Prob > F        =    0.0000
    Residual |  337845.354        85  3974.65122   R-squared       =    0.6319
-------------+----------------------------------   Adj R-squared   =    0.6233
       Total |  917854.506        87  10550.0518   Root MSE        =    63.045

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .1284362   .0138245     9.29   0.000     .1009495    .1559229
       bdrms |   15.19819   9.483517     1.60   0.113    -3.657582    34.05396
       _cons |    -19.315   31.04662    -0.62   0.536    -81.04399      42.414
------------------------------------------------------------------------------

.
. reg sqrft bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(1, 86)        =     33.85
       Model |  8186960.43         1  8186960.43   Prob > F        =    0.0000
    Residual |  20797100.3        86  241826.748   R-squared       =    0.2825
-------------+----------------------------------   Adj R-squared   =    0.2741
       Total |  28984060.7        87  333150.123   Root MSE        =    491.76

------------------------------------------------------------------------------
       sqrft | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       bdrms |   364.5886   62.6605     5.82   0.000     240.0236    489.1535
       _cons |   712.7749   229.6473     3.10   0.003     256.2513     1169.299
------------------------------------------------------------------------------
```

```
. predict e, resid

.
. reg price e

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(1, 86)        =     51.33
       Model |  343066.064         1  343066.064   Prob > F        =    0.0000
    Residual |  574788.441        86  6683.58653   R-squared       =    0.3738
-------------+----------------------------------   Adj R-squared   =    0.3665
       Total |  917854.506        87  10550.0518   Root MSE        =    81.753

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           e |   .1284362   .0179268     7.16   0.000     .0927989    .1640736
       _cons |    293.546   8.714921    33.68   0.000     276.2213    310.8707
------------------------------------------------------------------------------
```

The coefficients of the residual's component e_i are equal to the $\hat{\beta}_1$ estimated in the first regression. By definition, the coefficient $\hat{\beta}_1$ is the partial effect of $SQRFT_i$ on $PRICE_i$ holding everything else fixed. With this procedure, estimating the residuals of $SQRFT_i$ on $BDRMS_i$, we account for the variations in $SQRFT_i$ which does not depend on the variable $SQRFT_i$. In other words, we keep fixed the variation of $SQRFT_i$ due to change in $BDRMS_i$, and using the residuals we just see the partial effect of $SQRFT_i$ on prices.

(b) Solution:

```
. /// 1.B
>
. reg bdrms sqrft

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(1, 86)        =     33.85
       Model |  17.3972288         1  17.3972288   Prob > F        =    0.0000
    Residual |  44.1936803        86  .513880004   R-squared       =    0.2825
-------------+----------------------------------   Adj R-squared   =    0.2741
       Total |  61.5909091        87  .707941484   Root MSE        =    .71685

------------------------------------------------------------------------------
       bdrms | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .0007747   .0001332     5.82   0.000       .00051    .0010394
       _cons |   2.008077   .2788063     7.20   0.000     1.453829    2.562326
------------------------------------------------------------------------------
```

```
. gen delta_1 = _b[sqrft]

.
. reg price sqrft

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(1, 86)        =    140.79
       Model |  569801.074         1  569801.074   Prob > F        =    0.0000
    Residual |  348053.432        86  4047.13293   R-squared       =    0.6208
-------------+----------------------------------   Adj R-squared   =    0.6164
       Total |  917854.506        87  10550.0518   Root MSE        =    63.617

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |    .140211   .0118166    11.87   0.000     .1167203    .1637017
       _cons |   11.20415   24.74261     0.45   0.652    -37.98253    60.39082
------------------------------------------------------------------------------

.
. reg price sqrft bdrms

      Source |       SS           df       MS      Number of obs   =        88
-------------+----------------------------------   F(2, 85)        =     72.96
       Model |  580009.152         2  290004.576   Prob > F        =    0.0000
    Residual |  337845.354        85  3974.65122   R-squared       =    0.6319
-------------+----------------------------------   Adj R-squared   =    0.6233
       Total |  917854.506        87  10550.0518   Root MSE        =    63.045

------------------------------------------------------------------------------
       price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       sqrft |   .1284362   .0138245     9.29   0.000     .1009495    .1559229
       bdrms |   15.19819   9.483517     1.60   0.113    -3.657582    34.05396
       _cons |    -19.315   31.04662    -0.62   0.536    -81.04399      42.414
------------------------------------------------------------------------------

. gen  beta_1_hat = _b[sqrft]

. gen      beta_2_hat = _b[bdrms]

.
. gen error = beta_1_tilde - beta_1_hat - beta_2_hat*delta_1
```

Results are presented in Table 1.

Table 1: Estimation results from above steps.

$\tilde{\delta}_1$	$\tilde{\beta}_1$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\tilde{\beta}_1 - \hat{\beta}_1 - \hat{\beta}_2\tilde{\delta}_1$
0.0007747	0.140211	0.1284362	15.19819	-3.23e-09

We can verify that:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$$

To derive the expectation result, we take the conditional expectation on both sides and use the fact that the **OLS estimator** is conditionally unbiased:

$$\mathbb{E}[\hat{\beta}_1] = \beta_1, \quad \mathbb{E}[\hat{\beta}_2] = \beta_2.$$

Furthermore, **$\tilde{\delta}_1$ is a function of the data matrix X**:

$$\tilde{\delta}_1 = (X_1'X_1)^{-1}X_1'X_2.$$

Thus:

$$\mathbb{E}[\hat{\beta}_1 \mid X] = \mathbb{E}[\hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1 \mid X] = \beta_1 + \beta_2\tilde{\delta}_1.$$

We consider a regression model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,$$

where **X_{i2} is omitted** from the regression. Define:

$$X_i = \begin{bmatrix} 1 \\ SQRFT_i \\ BDRMS_i \end{bmatrix}, \quad Z_{i1} = \begin{bmatrix} 1 \\ SQRFT_i \end{bmatrix}, \quad Z_{i2} = \begin{bmatrix} BDRMS_i \end{bmatrix}.$$

Let:

$$\gamma_1 = (\beta_0, \beta_1)', \quad \gamma_2 = (\beta_2).$$

$$\hat{\beta} = (Z_1'Z_1)^{-1}Z_1'Y.$$

Substituting $Y = Z\beta + \varepsilon$:

$$\hat{\beta} = (Z_1'Z_1)^{-1}Z_1'(Z\beta + \varepsilon).$$

Expanding:

$$\hat{\beta} = \gamma_1 + P_{1.2}\gamma_2 + (Z_1'Z_1)^{-1}Z_1'\varepsilon.$$

Taking Expectation:

$$\mathbb{E}[\hat{\beta} \mid X] = \gamma_1 + P_{1.2}\gamma_2.$$

Since Z_2 is a scalar, $\gamma_2 = \beta_2$, and:

$$P_{1.2} = (\tilde{\delta}_0, \tilde{\delta}_1)'.$$

Thus, we obtain:

$$\mathbb{E}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \Bigg| X = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \tilde{\delta}_0 \\ \tilde{\delta}_1 \end{bmatrix} \beta_2.$$

The second row confirms the required result.

2 Question 2

Solution:

P_2 :

(a) When $L \times k$, $\hat\beta_{IV} = (Z'X)^{-1}Z'y = (Z'X)^{-1}Z'(X\beta+\varepsilon) = (Z'X)^{-1}Z'X\beta + (Z'X)^{-1}Z'\varepsilon$

$$= \beta + (Z'X)^{-1}Z'\varepsilon. = \beta + (\sum_{i=1}^{n} z_i x_i')^{-1}(\sum_{i=1}^{n} z_i'\varepsilon)$$

$$\Rightarrow \beta + (\sum_{i=1}^{n}(z_i x_i')^{-1}(\sum_{i=1}^{n} z_i'\varepsilon) \quad \to_p E z_i x_i \quad E z_i\varepsilon_i)$$

$$\to_p \beta$$

$\hat\beta_{IV} - \beta = (\sum_{i=1}^{n} z_i x_i')^{-1}(\sum_{i=1}^{n} z_i'\varepsilon)$, $\hat\beta_{2SLS} = (Z'X)^{-1}Z'y$, Let $P_Z = Z(Z'Z)^{-1}Z'$

$$\hat X = P_Z X = Z(Z'Z)^{-1}Z' \cdot X$$

$$\hat\beta_{2SLS} = (X'P_Z X)^{-1}X'P_Z y = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Xy$$

$$\hat\beta_{2SLS} - \beta = [X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon$$

(b) Moment of x_i and z_i, $E[x_i x_i'] = Q_{xx} < \infty$, rank $Q_{xx} = k$

$E[z_i z_i'] = Q_{zz} < \infty$. rank $Q_{zz} = L$

$E[z_i x_i'] = \underset{(L\times k)}{Q_{zx}} < \infty$ rank $Q_{zx} = k$.

By CLT and WLLN,

$$\sqrt{n}(\hat\beta_{2SLS} - \beta) = \sqrt{n}\underbrace{\left[(\tfrac{1}{n}\sum_{i=1}^{n} x_i z_i')(\tfrac{1}{n}\sum_{i=1}^{n} z_i z_i')^{-1}(\tfrac{1}{n}\sum_{i=1}^{n} z_i x_i')\right]^{-1}(\tfrac{1}{n}x_i z_i')(\tfrac{1}{n}\sum_{i=1}^{n} z_i z_i')^{-1}}(\sqrt{n}\,\tfrac{1}{n}\sum_{i=1}^{n} z_i\varepsilon_i)$$

$$\to_p (Q_{xz} \cdot Q_{zz}^{-1} \cdot Q_{zx})^{-1} \cdot Q_{xz} \cdot Q_{zz}^{-1}.$$

$$\sqrt{n} \cdot \frac{Z'\varepsilon}{n} = \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^{n} z_i \varepsilon_i \to_d N(0, \Omega)$$

$$\sqrt{n}(\hat\beta_{2SLS} - \beta) \to_d (Q_{xz} \cdot Q_{zz}^{-1} Q_{zx})^{-1} Q_{xz} Q_{zz}^{-1} \cdot N(0, \Omega). = N(0, V)$$

$$V = (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1} Q_{xz}(Q_{zz}^{-1}\Omega Q_{zz}^{-1} Q_{zx})(Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}$$

c. i) If under Homoskedasticity, $E(\varepsilon_i^2) = \sigma^2$

Then $\left(\sqrt{n}\,\frac{1}{n}\sum z_i \varepsilon_i\right) \xrightarrow{d} N(0, \sigma^2 Q_{zz})$

~~V can red~~ $\Rightarrow V = \sigma^2 (Q_{xz} Q_{zz}^{-1} Q_{zx})^{-1}$

A sample analog $\hat{V} = \hat{\sigma}^2 (\hat{Q}_{xz} \hat{Q}_{zz} \hat{Q}_{zx})^{-1}$ where

$\hat{Q}_{xz} = \frac{1}{n} X'Z$; $\hat{Q}_{zz} = \frac{1}{n} Z'Z$; $\hat{Q}_{zx} = \frac{1}{n} Z'X$

$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i'\hat{\beta}_{2SLS})^2$

ii) If under heteroskedasticity, $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' \hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' (y_i - x_i'\hat{\beta}_{2SLS})^2$

$\neq \frac{1}{n}\sum \hat{\varepsilon}_i^2$

d. (i) Under Homoskedasticity, $\hat{V} = \hat{\sigma}^2 (\hat{Q}_{xz} \hat{Q}_{zz} \hat{Q}_{zx})^{-1}$

$\hat{\varepsilon}_i = y_i - x'\hat{\beta} = x_i'\beta + \varepsilon_i - x_i'\hat{\beta} = x_i'(\beta - \hat{\beta}) + \varepsilon_i$

$\hat{\varepsilon}_i^2 = (y_i - x_i'\hat{\beta})^2 = (x_i'(\beta - \hat{\beta}) + \varepsilon_i)^2 = (\beta - \hat{\beta})' x_i x_i' (\beta - \hat{\beta}) + 2(\beta - \hat{\beta})' x_i \varepsilon_i + \varepsilon_i^2$

$\hat{\sigma}_i^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 + 2(\beta - \hat{\beta})' \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i + (\beta - \hat{\beta})' \frac{1}{n}\sum_{i=1}^{n} x_i x_i' (\beta - \hat{\beta})$

By WLLN, $\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 \xrightarrow{p} \sigma^2$;

Since $E x_{ik}^2 < \infty$ and $E\varepsilon_i^2 < \infty$, $E x_{ik}\varepsilon_i < \infty$, we can get

$\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i \xrightarrow{p} 0$ (By WLLN)

By WLLN, $\frac{1}{n}\sum_{i=1}^{n} x_i' x_i \xrightarrow{p} E x_i' x_i = 0$ as $n \to \infty$

Thus: $\hat{\sigma} \xrightarrow{p} \sigma$. $\hat{V} \xrightarrow{p} V$.

(ii) If under heteroskedasticity,

$\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} z_i z_i' (y_i - x_i'\hat{\beta}_{2SLS})^2$

$= \frac{1}{n}\sum z_i z_i' \varepsilon_i^2 + 2 \cdot \frac{1}{n}\sum_{i=1}^{n} z_i z_i' [(\beta - \hat{\beta})' x_i \varepsilon_i] + \frac{1}{n}\sum_{i=1}^{n} z_i z_i' [(\beta - \hat{\beta})' x_i x_i' (\beta - \hat{\beta})]$

$\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \varepsilon_i^2 \xrightarrow{p} E[z_i z_i' \varepsilon_i^2] < \infty$

$\Omega_{kl} = \frac{1}{n}\sum z_{ik} z_{il} [(\beta - \hat{\beta})' x_i \varepsilon_i] = (\beta - \hat{\beta})' \frac{1}{n}\sum z_{ik} z_{il} x_i \varepsilon_i$

$\xrightarrow{p} E[z_{ik} z_{il} x_i \varepsilon_i] < \infty$.

3 Question 3

Solution:

```
.
. // A
. eststo: ivregress 2sls lwage exper tenure (educ = sibs feduc meduc), first

First-stage regressions
-----------------------

                                                Number of obs =     722
                                                F(5, 716)     =   73.49
                                                Prob > F      =  0.0000
                                                R-squared     =  0.3391
                                                Adj R-squared =  0.3345
                                                Root MSE      =  1.8247

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       exper |  -.2009287   .0173033   -11.61   0.000    -.2348999   -.1669575
      tenure |    .037542   .0141963     2.64   0.008     .0096706    .0654135
        sibs |  -.1085301   .0317095    -3.42   0.001    -.1707849   -.0462754
       feduc |   .1568147   .0256903     6.10   0.000     .1063774    .2072521
       meduc |    .110667    .030085     3.68   0.000     .0516017    .1697323
       _cons |   13.17707   .4193578    31.42   0.000     12.35375    14.00039
------------------------------------------------------------------------------


Instrumental variables 2SLS regression            Number of obs   =        722
                                                  Wald chi2(3)    =      76.58
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.0641
                                                  Root MSE        =     .40543

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .1376307   .0183256     7.51   0.000     .1017133    .1735482
       exper |    .033928   .0058711     5.78   0.000     .0224208    .0454351
      tenure |   .0073686   .0032532     2.27   0.024     .0009925    .0137447
       _cons |   4.481904   .2992201    14.98   0.000     3.895443    5.068365
------------------------------------------------------------------------------
. // B
. eststo: reg educ exper tenure sibs feduc meduc

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(5, 716)       =     73.49
       Model |  1223.3278         5  244.665559   Prob > F        =    0.0000
    Residual |  2383.88689       716  3.32945096   R-squared       =    0.3391
-------------+----------------------------------   Adj R-squared   =    0.3345
       Total |  3607.21468       721  5.00307168   Root MSE        =    1.8247

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       exper |  -.2009287   .0173033   -11.61   0.000    -.2348999   -.1669575
      tenure |    .037542   .0141963     2.64   0.008     .0096706    .0654135
        sibs |  -.1085301   .0317095    -3.42   0.001    -.1707849   -.0462754
       feduc |   .1568147   .0256903     6.10   0.000     .1063774    .2072521
       meduc |    .110667    .030085     3.68   0.000     .0516017    .1697323
       _cons |   13.17707   .4193578    31.42   0.000     12.35375    14.00039
------------------------------------------------------------------------------

(est2 stored)
```

```
. predict educ_hat
(option xb assumed; fitted values)
(213 missing values generated)

. eststo: reg lwage educ_hat exper tenure

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     26.37
       Model |  12.5876366         3  4.19587887   Prob > F        =    0.0000
    Residual |  114.224295       718  .159086761   R-squared       =    0.0993
-------------+----------------------------------   Adj R-squared   =    0.0955
       Total |  126.811931       721    .1758834   Root MSE        =    .39886

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    educ_hat |   .1376307   .0180283     7.63   0.000     .1022362    .1730252
       exper |    .033928   .0057759     5.87   0.000     .0225884    .0452675
      tenure |   .0073686   .0032004     2.30   0.022     .0010853    .0136518
       _cons |   4.481904   .2943666    15.23   0.000     3.903982    5.059826
------------------------------------------------------------------------------
(est3 stored)

. // For the Table
. reg educ exper tenure sibs feduc meduc

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(5, 716)       =     73.49
       Model |   1223.3278         5  244.665559   Prob > F        =    0.0000
    Residual |  2383.88689       716  3.32945096   R-squared       =    0.3391
-------------+----------------------------------   Adj R-squared   =    0.3345
       Total |  3607.21468       721  5.00307168   Root MSE        =    1.8247

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       exper |  -.2009287   .0173033   -11.61   0.000    -.2348999   -.1669575
      tenure |    .037542   .0141963     2.64   0.008     .0096706    .0654135
        sibs |  -.1085301   .0317095    -3.42   0.001    -.1707849   -.0462754
       feduc |   .1568147   .0256903     6.10   0.000     .1063774    .2072521
       meduc |    .110667    .030085     3.68   0.000     .0516017    .1697323
       _cons |   13.17707   .4193578    31.42   0.000     12.35375    14.00039
------------------------------------------------------------------------------

. reg lwage educ_hat exper tenure

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     26.37
       Model |  12.5876366         3  4.19587887   Prob > F        =    0.0000
    Residual |  114.224295       718  .159086761   R-squared       =    0.0993
-------------+----------------------------------   Adj R-squared   =    0.0955
       Total |  126.811931       721    .1758834   Root MSE        =    .39886

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    educ_hat |   .1376307   .0180283     7.63   0.000     .1022362    .1730252
       exper |    .033928   .0057759     5.87   0.000     .0225884    .0452675
      tenure |   .0073686   .0032004     2.30   0.022     .0010853    .0136518
       _cons |   4.481904   .2943666    15.23   0.000     3.903982    5.059826
------------------------------------------------------------------------------
```

We can notice that all estimates β_j are identical to those obtained before, but that the standard errors in the second equation are different. In the second case, the estimate of σ^2 is wrong. Its computation includes the variance of the sum of the error term in the second equation and the error term from the first stage regression. Since $\hat{X}$ is a projection of X, 2SLS will have higher variance (even if OLS is inconsistent). This confirms the fact that, when OLS is consistent, it is better than 2SLS. But when X is endogenous, we must account for endogeneity using the 2SLS procedure. In summary, we can conclude that standard errors from the second approach are invalid.

```
. // c
. eststo: reg educ  sibs feduc meduc

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     65.20
       Model |  772.281437         3  257.427146   Prob > F        =    0.0000
    Residual |  2834.93324       718  3.94837499   R-squared       =    0.2141
-------------+----------------------------------   Adj R-squared   =    0.2108
       Total |  3607.21468       721  5.00307168   Root MSE        =    1.9871

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        sibs |  -.0936359   .0344713    -2.72   0.007    -.1613124   -.0259594
       feduc |   .2100041   .0274748     7.64   0.000     .1560635    .2639447
       meduc |   .1307872    .032689     4.00   0.000     .0666098    .1949646
       _cons |   10.36426   .3585001    28.91   0.000     9.660422    11.06809
------------------------------------------------------------------------------
(est4 stored)

. predict educ_tilde
(option xb assumed; fitted values)
(213 missing values generated)

. eststo: reg lwage educ_tilde exper tenure

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     26.45
       Model |   12.619224         3  4.20640799   Prob > F        =    0.0000
    Residual |  114.192707       718  .159042768   R-squared       =    0.0995
-------------+----------------------------------   Adj R-squared   =    0.0957
       Total |  126.811931       721    .1758834   Root MSE        =     .3988

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  educ_tilde |   .1129465   .0147677     7.65   0.000     .0839534    .1419396
       exper |   .0065613   .0037657     1.74   0.082    -.0008319    .0139544
      tenure |   .0126207   .0030986     4.07   0.000     .0065374     .018704
       _cons |   5.091582   .2153205    23.65   0.000     4.668849    5.514315
------------------------------------------------------------------------------
(est5 stored)
```

```
.
. // For the Table
. reg educ  sibs feduc meduc

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     65.20
       Model |  772.281437         3  257.427146   Prob > F        =    0.0000
    Residual |  2834.93324       718  3.94837499   R-squared       =    0.2141
-------------+----------------------------------   Adj R-squared   =    0.2108
       Total |  3607.21468       721  5.00307168   Root MSE        =    1.9871

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        sibs |  -.0936359   .0344713    -2.72   0.007    -.1613124   -.0259594
       feduc |   .2100041   .0274748     7.64   0.000     .1560635    .2639447
       meduc |   .1307872    .032689     4.00   0.000     .0666098    .1949646
       _cons |   10.36426   .3585001    28.91   0.000     9.660422    11.06809
------------------------------------------------------------------------------

. reg lwage educ_tilde exper tenure

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(3, 718)       =     26.45
       Model |   12.619224         3  4.20640799   Prob > F        =    0.0000
    Residual |  114.192707       718  .159042768   R-squared       =    0.0995
-------------+----------------------------------   Adj R-squared   =    0.0957
       Total |  126.811931       721    .1758834   Root MSE        =     .3988

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  educ_tilde |   .1129465   .0147677     7.65   0.000     .0839534    .1419396
       exper |   .0065613   .0037657     1.74   0.082    -.0008319    .0139544
      tenure |   .0126207   .0030986     4.07   0.000     .0065374     .018704
       _cons |   5.091582   .2153205    23.65   0.000     4.668849    5.514315
------------------------------------------------------------------------------
```

The first stage is used to estimate the endogenous variable, and if the first stage is not correctly specified, the second stage will not be consistent. The error term in the second stage will not be orthogonal to the exogenous regressors and the estimator will be inconsistent.

```
. // D
. eststo: reg lwage educ exper tenure

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(3, 931)       =     56.97
       Model |  25.6953278         3  8.56510927   Prob > F        =    0.0000
    Residual |  139.960966       931  .150334013   R-squared       =    0.1551
-------------+----------------------------------   Adj R-squared   =    0.1524
       Total |  165.656294       934  .177362199   Root MSE        =    .38773

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .0748638   .0065124    11.50   0.000      .062083    .0876446
       exper |   .0153285   .0033696     4.55   0.000     .0087156    .0219413
      tenure |   .0133748   .0025872     5.17   0.000     .0082974    .0184522
       _cons |   5.496696   .1105282    49.73   0.000     5.279782    5.713609
------------------------------------------------------------------------------
(est6 stored)
```

```
. esttab, se ar2 replace

--------------------------------------------------------------------------------
> ----------------------------
                      (1)             (2)             (3)             (4)
>             (5)             (6)
                    lwage            educ           lwage            educ
>           lwage           lwage
--------------------------------------------------------------------------------
> ----------------------------
educ                0.138***
>                       0.0749***
                 (0.0183)
>                     (0.00651)

exper              0.0339***       -0.201***       0.0339***
>       0.00656       0.0153***
                (0.00587)        (0.0173)       (0.00578)
> (0.00377)     (0.00337)

tenure            0.00737*         0.0375**       0.00737*
>      0.0126***      0.0134***
                (0.00325)        (0.0142)       (0.00320)
> (0.00310)     (0.00259)

sibs                               -0.109***                       -0.0936**
>
                                 (0.0317)                        (0.0345)
>

feduc                               0.157***                        0.210***
>
                                 (0.0257)                        (0.0275)
>

meduc                               0.111***                        0.131***
>
                                 (0.0301)                        (0.0327)
>

educ_hat                                            0.138***
>
                                                 (0.0180)
>

educ_tilde
>        0.113***

>  (0.0148)

_cons               4.482***        13.18***        4.482***        10.36***
>        5.092***        5.497***
                  (0.299)         (0.419)         (0.294)         (0.359)
>   (0.215)         (0.111)
--------------------------------------------------------------------------------
> ----------------------------
N                     722             722             722             722
>             722             935
adj. R-sq           0.060           0.335           0.095           0.211
>           0.096           0.152
--------------------------------------------------------------------------------
```

4 Question 4

Solution:

(4) From the conclusion in Prob (2), we have

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y$$

$$= \left(\begin{bmatrix} 219.4 & 115.4 & 49.6 \\ 52.4 & 34.6 & 61.4 \end{bmatrix} \cdot \begin{bmatrix} 526 & 118.7 & 59.2 \\ 188.7 & 211.9 & 42.9 \\ 59.2 & 42.9 & 194.7 \end{bmatrix} \begin{bmatrix} 219 & 52.4 \\ 115.4 & 34.6 \\ 49.6 & 61.4 \end{bmatrix} \right)^{-1} \cdots$$

$$= \begin{pmatrix} 1.728 \\ 1.654 \end{pmatrix} \quad \text{(By MATLab)}$$

(b) $$\hat{\sigma}^2 = \frac{e'e}{n} = \frac{Y'Y + \hat{\beta}_{2SLS}' X'X \hat{\beta}_{2SLS} - 2(X\hat{\beta}_{2SLS})'Y}{n} = 7.788$$

$$\hat{\sigma}^2 [X'Z(Z'Z)^{-1}Z'X]^{-1} = \begin{bmatrix} 0.13328 & -0.1884 \\ -0.1884 & 0.6168 \end{bmatrix}$$

t-statistic: $\dfrac{\hat{\beta}_{2SLS,1} - \beta_1}{se(\hat{\beta}_{2SLS})} = \dfrac{1.7279-1}{\sqrt{0.13328}} = 1.9939$, By ~~took~~ checking with T-table, the p-value is 0.023

$\Rightarrow$ we can reject H_0 at 5% level.

(c) $\Rightarrow CI = (1.6535 - 1.96 \cdot \sqrt{0.6168},\ 1.6535 + 1.96 \cdot \sqrt{0.6168})$

(d) $H_0: \beta_1 = 1$ and $\beta_2 = 2$. under H_0.

$$\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = \begin{pmatrix} 1.728 - 1 \\ 1.654 - 2 \end{pmatrix} = \begin{pmatrix} 0.728 \\ -0.347 \end{pmatrix}, \quad var\begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{pmatrix} = \begin{pmatrix} 0.1333 & -0.1884 \\ -0.1884 & 0.6168 \end{pmatrix}$$

By using Wald test,

$(\hat{\beta} - \beta)' Var(\hat{\beta})^{-1} (\hat{\beta} - \beta) = 5.3056$, p-value $= 0.07045 > 0.01$

reject H_0. at 1%

5 Question 5

Solution:

P5. $Y_i = Z_i'\beta + \eta_i$

(a). First, we regress z_i on W_i (IV) to get $\hat{Z}_i$

$$Z_i = X_i = \delta_0 + \delta_1 W_i + u_i$$

Since $W_i = X_i + v_i$, $Z_i = X_i = \delta_0 + \delta_1 (X_i + v_i) + u_i$

$$\hat{\delta}_1 = \frac{Cov(Z_i, W_i)}{Var(W_i)} = \frac{Cov(X_i, X_i + v_i)}{Var(X_i + v_i)} = \frac{Var X_i + 0}{Var X_i + \sigma_v^2} < 1$$

(b) In second stage. we regress Y_i on $\hat{Z}_i$ which $\hat{z}_i = \hat{\delta}_0 + \hat{\delta}_1 w_i$, And under homoskedacity assumption, $\sqrt{n}(\hat{\beta}_{IV} - \beta) \to_d N(0, V_{IV})$

$$V_{IV} = \sigma_\varepsilon^2 E[W_i Z_i']^{-1} E[W_i W_i'] E[Z_i W_i']^{-1}$$

$$E[W_i Z_i'] = E[(X_i + v_i) X_i] = E X_i^2 + E v_i X_i = E X_i^2$$

$$E[W_i W_i'] = E[(X_i + v_i)^2] = E[X_i^2] + 2E[X_i v_i] + E v_i^2 = E[X_i^2] + \sigma_v^2$$

$$\therefore V_{IV} = \sigma_\varepsilon^2 \cdot \frac{1}{E[X_i^2]} \{E[X_i^2] + \sigma_v^2\} \cdot \frac{1}{E[X_i^2]} = \sigma_\varepsilon^2 \cdot \frac{E[X_i^2] + \sigma_v^2}{E[X_i^2]^2}$$

(c). From (b). when $\sigma_v^2 = 0$, V_{IV} have same limiting distribution as OLS estimation.