황 계량 (2020기기기)

1. Given

$$Y = X\beta + U,$$

suppose one or some of explanatory variables are endogenous.
1) What happens to the OLS estimator $\hat{\beta}_{OLS}$?
2) You may want to use IV. Show that the IV estimator of β is consistent.
3) Choose the best IV and compare it with the two-stage least squares estimator of β.
4) Obtain the asymptotic distribution of the two-stage least squares estimator of β.

1) endogenous indep var : $E[U|X] \neq 0$, A3 violated

$\hat{\beta}_{OLS}$ biased and inconsistent

2) def of IV,

① $\frac{1}{n} Z'U \xrightarrow{p} 0$

② $\frac{1}{n} Z'X \xrightarrow{p} Q_{ZX}$ of full col. rank.

$\hat{\beta}_{IV}$ solves $\frac{1}{n} Z'(Y - X\beta) = 0$.

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'Y = \beta + (Z'X)^{-1} Z'U \xrightarrow{p} \beta + 0 \text{ by ①}$$

$\therefore \hat{\beta}_{IV} \xrightarrow{p} \beta$ for $n \gg 1$, consistent

3) $\hat{\beta}_{best\ IV} = (\hat{X}'X)^{-1} \hat{X}'Y$, where $\hat{X} = Z(Z'Z)^{-1} Z'X$

(2SLS): [1st stage] reg X on Z

$X = Z\alpha + U$, $\hat{\alpha} = (Z'Z)^{-1} Z'X$ $\therefore \hat{X} = Z(Z'Z)^{-1} Z'X$

[2nd stage] reg Y on $\hat{X}$

$Y = \hat{X}\beta + U$, $\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1} \hat{X}'Y = (\hat{X}'X)^{-1} \hat{X}'Y = \hat{\beta}_{best\ IV}$. same

4) since $\hat{\beta}_{2SLS} = \hat{\beta}_{best\ IV}$, find asy. dist'n of $\hat{\beta}_{best\ IV}$

recall $\hat{\beta}_{best\ IV} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y$, where $\hat{X} = Z(Z'Z)^{-1}Z'X$

$$\hat{\beta} = \beta + (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'u$$

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}X'Z\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{n}Z'X\right)^{-1}\frac{1}{n}X'Z\left(\frac{1}{n}Z'Z\right)^{-1}\frac{1}{\sqrt{n}}Z'u$$

$$\xrightarrow{d} (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q_{XZ}Q_{ZZ}^{-1}N(0, \sigma^2 Q_{ZZ})$$

$$= N(0, \sigma^2 (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q_{XZ}Q_{ZZ}^{-1}Q_{ZZ}Q_{ZZ}^{-1}Q_{ZX}(\cdot)^{-1})$$

$$= N(0, \sigma^2 (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1})$$

$$\therefore \hat{\beta}_{best\ IV} = \hat{\beta}_{2SLS} \xrightarrow{d} N\left(\beta, \frac{\sigma^2}{n}(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}\right)$$

2.

1) Results of OLS estimation are given as follows. OLS regression suggests that the es-

```
t test of coefficients:

              Estimate Std. Error  t value  Pr(>|t|)    
(Intercept) -4.1656505  0.2409182 -17.2907 < 2.2e-16 ***
educ        -0.0899641  0.0059230 -15.1889 < 2.2e-16 ***
age          0.3344319  0.0165813  20.1693 < 2.2e-16 ***
age2        -0.0026708  0.0002733  -9.7723 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

timated effect for 1 more year of education is a reduction of the number of children by 0.09. Loosely speaking, if 100 women were to receive another year of education, one may suggest that approximately 9 fewer children should be expected.

Moreover, note that the coefficients on `age` and `age2` are both significant, suggesting that the effect of age on number of children takes a quadratic form, perhaps due to early deaths of male descendants.

2) `educ` is regressed on `frsthalf` to see if the two variables have a non-zero covariance, suggesting that `frsthalt` is a resonable IV candidate. Results are as follows.

```
t test of coefficients:

               Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)  9.69778205  0.59929703 16.1819 < 2.2e-16 ***
frsthalf    -0.85210325  0.11291893 -7.5462 5.428e-14 ***
age         -0.10853136  0.04215247 -2.5747   0.01006 *  
age2        -0.00048301  0.00069527 -0.6947   0.48727    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The coefficient of `frsthalf` is significant, suggesting that `frsthalf` is correlated with `educ`, satisfying one of the two conditions of a reasonable IV.

3) The following results estimates the model from 1) using `frsthalf` as an IV for `educ`, as `frsthalf` satisfies both conditions of a reasonable IV.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.4434082  0.5479168  -6.285 3.61e-10 ***
educ        -0.1677795  0.0531121  -3.159  0.00159 ** 
age          0.3258729  0.0178757  18.230  < 2e-16 ***
age2        -0.0027087  0.0002799  -9.679  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note that the effect of education has increased twice-fold while still being statistically significant. That is, under IV estimation, results imply that an additional year of education corresponds to a 0.17 decrease in the number of children.

However, given the std. error of `educ(IV)`, we can approximate the 95% confidence interval, and see that the interval(-0.06 -0.26) includes the OLS estimate on `educ`. In this sense, it may be difficult to suggest that the coefficients from both estimations are statistically different.

4) `tv` is included in the model, and the OLS estimates (using `educ`) and 2SLS estimates (using `frsthalf` as an IV for `educ`) are given as follows.

```
t test of coefficients:

              Estimate   Std. Error  t value  Pr(>|t|)
(Intercept) -4.24908177  0.24097474 -17.6329 < 2.2e-16 ***
educ        -0.08003108  0.00627067 -12.7628 < 2.2e-16 ***
age          0.33691359  0.01654897  20.3586 < 2.2e-16 ***
age2        -0.00267963  0.00027264  -9.8286 < 2.2e-16 ***
tv          -0.38137428  0.08063589  -4.7296 2.321e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
t test of coefficients:

              Estimate   Std. Error t value  Pr(>|t|)
(Intercept) -3.44461428  0.62011468 -5.5548 2.945e-08 ***
educ        -0.16763950  0.06230604 -2.6906   0.00716 **
age          0.32590303  0.01862358 17.4995 < 2.2e-16 ***
age2        -0.00270881  0.00027946 -9.6932 < 2.2e-16 ***
tv          -0.00405716  0.27935673 -0.0145   0.98841
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Looking at the OLS estimates, one may claim that television ownership has a statistically significant negative effect on fertility. However, given that `educ` is endogenous, these results are inconsistent with the true model. Implementing `frsthalf` as an IV to solve for this problem, it can be seen that the negative effect of TV ownership is no longer statistically significant.