HW0428_313707006_王彩寧

10.18

(a)

```
> cat(round(mother_pct, 2), "%\n")
12.15 %
> cat(round(father_pct, 2), "%\n")
11.68 %
```

(b)

```
> print(round(cor_matrix, 3))
            educ mothercoll fathercoll
educ       1.000      0.359      0.398
mothercoll 0.359      1.000      0.355
fathercoll 0.398      0.355      1.000
```

The correlation between EDUC and MOTHERCOLL is 0.3595. The correlation between EDUC and FATHERCOLL is 0.3985. It might be argued that parents who had attended college for at least some time have a higher appreciation of education for their children than those who did not.

(c)

95% CI for EDUC coefficient: [-0.0012 , 0.1533]

(d)

```
Call:
lm(formula = educ ~ exper + exper2 + mothercoll, data = mroz2)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.4267 -0.4826 -0.3731  1.0000  4.9353 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 12.079094   0.303118  39.849  < 2e-16 ***
exper        0.056230   0.042101   1.336    0.182    
exper2      -0.001956   0.001256  -1.557    0.120    
mothercoll   2.517068   0.315713   7.973 1.46e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.133 on 424 degrees of freedom
Multiple R-squared:  0.1347,	Adjusted R-squared:  0.1285 
F-statistic: 21.99 on 3 and 424 DF,  p-value: 2.965e-13
```

```
Linear hypothesis test:
mothercoll = 0

Model 1: restricted model
Model 2: educ ~ exper + exper2 + mothercoll

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)    
1    425 2219.2                                  
2    424 1929.9  1    289.32 63.563 1.455e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The t-statistic for the coefficient of MOTHERCOLL is 7.97, and the corresponding F-value is 63.56. This is far greater than the rule-of-thumb value of 10, so we reject the notion that the IV is weak on this basis.

(e)

95% CI for EDUC (2 IVs): [0.0275 , 0.1482]

This is slightly narrower than the interval estimate using only MOTHERCOLL as an IV as shown in part (c).

(f)

```
Call:
lm(formula = educ ~ exper + exper2 + mothercoll + fathercoll,
    data = mroz2)

Residuals:
    Min      1Q  Median      3Q     Max
-7.2152 -0.3056 -0.2152  0.7627  5.0620

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.890259   0.290251  40.965  < 2e-16 ***
exper        0.049149   0.040133   1.225    0.221
exper2      -0.001449   0.001199  -1.209    0.227
mothercoll   1.749947   0.322347   5.429 9.58e-08 ***
fathercoll   2.186612   0.329917   6.628 1.04e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.033 on 423 degrees of freedom
Multiple R-squared:  0.2161,	Adjusted R-squared:  0.2086
F-statistic: 29.15 on 4 and 423 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
mothercoll = 0
fathercoll = 0

Model 1: restricted model
Model 2: educ ~ exper + exper2 + mothercoll + fathercoll

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    425 2219.2
2    423 1748.3  2    470.88 56.963 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We reject the null hypothesis that the instruments are weak.

(g)

```
> (p_value <- 1 - pchisq(S, df = 1))
[1] 0.6281333
```

we fail to reject the null hypothesis.

10.20

(a)

```
Call:
lm(formula = msft_excess ~ mkt_excess, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27424 -0.04744 -0.00820  0.03869  0.35801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003250   0.006036   0.538    0.591
mkt_excess  1.201840   0.122152   9.839   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08083 on 178 degrees of freedom
Multiple R-squared:  0.3523,	Adjusted R-squared:  0.3486
F-statistic:  96.8 on 1 and 178 DF,  p-value: < 2.2e-16
```

Microsoft beta is 1.2018.

Microsoft stock is relatively risky compared to the market portfolio.

(b)

```
Call:
lm(formula = mkt_excess ~ RANK, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.110497 -0.006308  0.001497  0.009433  0.029513 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -7.903e-02  2.195e-03   -36.0   <2e-16 ***
RANK         9.067e-04  2.104e-05    43.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01467 on 178 degrees of freedom
Multiple R-squared:  0.9126,	Adjusted R-squared:  0.9121 
F-statistic:  1858 on 1 and 178 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
RANK = 0

Model 1: restricted model
Model 2: mkt_excess ~ RANK

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)    
1    179 0.43784                                  
2    178 0.03829  1   0.39955 1857.6 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variable RANK does not directly cause Microsoft's return, so it meets condition IV1.

R^2 = 0.9126.

The t-value is 43.10, and the F-statistic is 1857.61.

This means that RANK is a very strong instrument.

(c)

```
Call:
lm(formula = msft_excess ~ mkt_excess + v_hat, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27140 -0.04213 -0.00911  0.03423  0.34887 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.003018   0.005984   0.504   0.6146    
mkt_excess   1.278318   0.126749  10.085   <2e-16 ***
v_hat       -0.874599   0.428626  -2.040   0.0428 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08012 on 177 degrees of freedom
Multiple R-squared:  0.3672,	Adjusted R-squared:   0.36 
F-statistic: 51.34 on 2 and 177 DF,  p-value: < 2.2e-16
```

The t-statistic on $\hat{v}$ is -2.04 with a p = 0.043.

It is not significant at the 1% level, but at the 5% level.

At the 1% level, we cannot reject the null hypothesis that the market return is exogenous.

(d)

```
Call:
ivreg(formula = msft_excess ~ mkt_excess | RANK, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.271625 -0.049675 -0.009693  0.037683  0.355579 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.003018   0.006044   0.499    0.618    
mkt_excess  1.278318   0.128011   9.986   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08092 on 178 degrees of freedom
Multiple R-Squared: 0.3508,	Adjusted R-squared: 0.3472 
Wald test: 99.72 on 1 and 178 DF,  p-value: < 2.2e-16 
```

The IV estimate of Microsoft's beta (1.2783) is slightly higher than the OLS estimate (1.2018), consistent with the expectation that OLS suffers from attenuation bias due to potential measurement error in the market excess return.

(e)

```
Call:
lm(formula = mkt_excess ~ RANK + POS, data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.109182 -0.006732  0.002858  0.008936  0.026652 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0804216  0.0022622  -35.55   <2e-16 ***
RANK         0.0009819  0.0000400   24.55   <2e-16 ***
POS         -0.0092762  0.0042156   -2.20   0.0291 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01451 on 177 degrees of freedom
Multiple R-squared:  0.9149,	Adjusted R-squared:  0.9139 
F-statistic: 951.3 on 2 and 177 DF,  p-value: < 2.2e-16
```

```
Linear hypothesis test:
RANK = 0
POS = 0

Model 1: restricted model
Model 2: mkt_excess ~ RANK + POS

  Res.Df     RSS Df Sum of Sq      F    Pr(>F)    
1    179 0.43784                                  
2    177 0.03727  2   0.40057 951.26 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the F-test, we would conclude the IVs are not weak.

(e)

```
Call:
lm(formula = msft_excess ~ mkt_excess + v_hat2, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27132 -0.04261 -0.00812  0.03343  0.34867 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.003004   0.005972   0.503   0.6157    
mkt_excess   1.283118   0.126344  10.156   <2e-16 ***
v_hat2      -0.954918   0.433062  -2.205   0.0287 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07996 on 177 degrees of freedom
Multiple R-squared:  0.3696,	Adjusted R-squared:  0.3625 
F-statistic: 51.88 on 2 and 177 DF,  p-value: < 2.2e-16
```

The t-value on the first stage residuals is -2.05 with a p-value of 0.0287. Thus, at the

1% level, we cannot reject the null hypothesis that the market return is exogenous.

(f)

```
Call:
ivreg(formula = msft_excess ~ mkt_excess | RANK + POS, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.27168 -0.04960 -0.00983  0.03762  0.35543

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003004   0.006044   0.497     0.62
mkt_excess  1.283118   0.127866  10.035   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08093 on 178 degrees of freedom
Multiple R-Squared: 0.3507,	Adjusted R-squared: 0.347
Wald test: 100.7 on 1 and 178 DF,  p-value: < 2.2e-16
```

The coefficient estimate 1.28 is larger than the OLS estimate from part (a). If there is a measurement error problem the OLS estimator suffers attenuation bias, that is, it is biased downward.

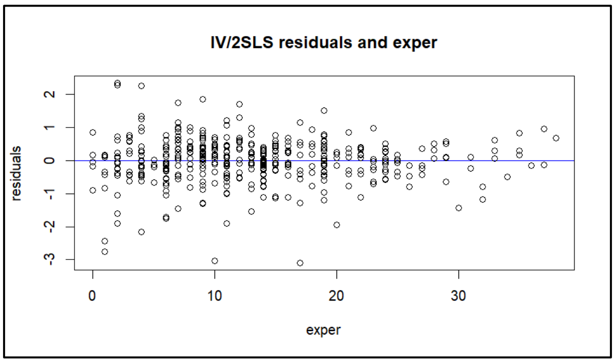
(h)

```
Call:
lm(formula = iv_resid ~ RANK + POS, data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max
-0.26914 -0.04702 -0.00801  0.03771  0.35674

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0022220  0.0126326  -0.176    0.861
RANK         0.0001370  0.0002234   0.613    0.540
POS         -0.0174499  0.0235409  -0.741    0.460

Residual standard error: 0.08103 on 177 degrees of freedom
Multiple R-squared:  0.003103,	Adjusted R-squared:  -0.008162
F-statistic: 0.2754 on 2 and 177 DF,  p-value: 0.7596

>
> R2 <- summary(sargan_reg)$r.squared
> n <- nrow(capm5)
> S <- n * R2
> S
[1] 0.5584634
> (p_value <- 1 - pchisq(S, df = 1))
[1] 0.45488
```

At the 5% level of significance, we find no evidence against the validity of RANK and POS as instruments, indicating that they can be treated as exogenous variables in the model.

10.24

(a)



There may be a “funnel” shape present, with more variation in the residuals at lower years of experience than at higher years of experience.

(b)

$nR^2 = 7.438552$ p-value = 0.006384122

We reject the null hypothesis of homoskedasticity.

(c)

```
> print(ci_robust)
       Lower        Upper
-0.003947005  0.126740261
```

```
> print(se_comparison)
            Estimate Baseline_SE Robust_SE Increased_SE
(Intercept)  0.04810     0.40033   0.42980          Yes
exper        0.04417     0.01343   0.01555          Yes
I(exper^2)  -0.00090     0.00040   0.00043          Yes
educ         0.06140     0.03144   0.03334          Yes
```

(d)

```
> print(boot_ci)
[1] -0.002000496  0.124793752
```

The bootstrap standard errors are ever so slightly smaller than the robust standard errors, but still a bit larger than the usual IV standard errors.