HW0428_313707002_羅芷羚

Q10.18:

(a)12.15% of mothers have some college education

11.68% of fathers have some college education.

```
> mean(mroz_sub$MOTHERCOLL, na.rm = TRUE)
[1] 0.1214953
> mean(mroz_sub$FATHERCOLL, na.rm = TRUE)
```

(B)The correlation between EDUC and MOTHERCOLL is 0.3595. The correlation between EDUC and FATHERCOLL is 0.3985.The education level from parents show positive correlation with education level for married women.

```
> cor(mroz_sub[, c("educ", "MOTHERCOLL", "FATHERCOLL")], use = "complete.obs")
                educ MOTHERCOLL FATHERCOLL
educ       1.0000000  0.3594705  0.3984962
MOTHERCOLL 0.3594705  1.0000000  0.3545709
FATHERCOLL 0.3984962  0.3545709  1.0000000
```

(c) The 95% confidence interval for one IV model

```
> confint(iv_model, level = 0.95)["educ", ]
       2.5 %       97.5 %
-0.001219763  0.153255678
```

(d) the first-stage equation for one IV

$$\hat{EDUC} = 12.0791 + 0.0562 \cdot EXPER - 0.00196 \cdot EXPER^2 + 2.5171 \cdot MOTHERCOLL + v$$

```
> summary(first_stage)

Call:
lm(formula = educ ~ exper + I(exper^2) + MOTHERCOLL, data = mroz_sub)

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    
I(exper^2)  -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
```

null hypothesis $H_0 : \beta_{MOTHERCOLL} = 0$

Alternative hypothesis $H_1 : \beta_{MOTHERCOLL} \neq 0$

The F value for MOTHERCOLL is 63.5631 larger than 10.We fail to reject H_0 ,this indicates that MOTHERCOLL has a significant effect on EDUC in the regression model. MOTHERCOLL is a strong instrumental variable.

```
> Anova(first_stage, type = "II")
Anova Table (Type II tests)

Response: educ
            Sum Sq  Df F value    Pr(>F)    
exper         8.12   1  1.7838    0.1824    
I(exper^2)   11.04   1  2.4254    0.1201    
mothercoll  289.32   1 63.5631 1.455e-14 ***
Residuals  1929.90 424                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(e) The 95% confidence interval for two IV model

```
> confint(iv_model_2, level = 0.95)["educ", ]
     2.5 %     97.5 %
0.02751845 0.14817686
```

The 95% confidence interval of model use two instrumental variable is narrower than the model use only one instrumental variable

	2.5%	97.5%
one IV model	−0.00122	0.15326
two IV model	0.027512	0.14818

(f) the first-stage equation for one IV

$$\hat{EDUC} = 11.89026 + 0.04915 \cdot EXPER - 0.00145 \cdot EXPER^2 + 1.74995 \cdot MOTHERCOLL$$
$$+2.18661 \cdot FATHERCOLL + \nu$$

```
Call:
lm(formula = educ ~ exper + exper2 + MOTHERCOLL + FATHERCOLL,
    data = mroz_sub)

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
```

The F-test statistic of the joint significance of MOTHERCOLL and FATHERCOLL is 56.96 >10, so we reject the null hypothesis that the instruments are weak.

```
> linearHypothesis(first_stage_2, c("MOTHERCOLL = 0", "FATHERCOLL = 0"))

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
```

(g) The p-value is 0.6281333, indicating that your instrument variables do not have any issues in the regression model and there are no overidentifying restrictions.

Q10.20:

(a) Based on this output, Microsoft stock appears to be risky relative to the market, as indicated by the positive and significant beta of 1.201840. The market return is a significant factor in explaining the excess return on Microsoft stock.

```
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
```

(b)The F-statistic is 1857.6, which is far greater than the common rule-of-thumb threshold of 10. This indicates that RANKis a strong instrument for market excess return.
Thus, based on the first-stage regression output, RANK can be regarded as a strong IV for the market return variable in the CAPM model.

```
> Anova(first_stage, type = "II")
Anova Table (Type II tests)

Response: mkt_excess
           Sum Sq  Df F value    Pr(>F)    
RANK      0.39955   1  1857.6 < 2.2e-16 ***
Residuals 0.03829 178                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c)The hypothesis that the market return is exogenous is not strongly supported at the 99% confidence level. However, there is some evidence of endogeneity at the 95% confidence level. Therefore, it's safe to conclude that the market return could be endogenous, but the evidence isn't overwhelmingly strong at the 1% significance level.

```
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
```

(d) The IV estimation gives a slightly higher coefficient for mkt_excess compared to the OLS model, which suggests that the IV model might be correcting for endogeneity, leading to a slightly stronger estimated relationship between the market excess return and Microsoft excess return.

```
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
```

(e)The F-test statistic for the joint significance of these variables is 951.26>10, indicating that RANK and pos are jointly significantly different from zero. Therefore, the null hypothesis that RANK and pos are both equal to zero can be rejected, suggesting that these variables are meaningful in the first stage of your regression.

```
> linearHypothesis(first_stage_pos, c("RANK = 0", "pos = 0"))

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
```

(f)At the 1% level we fail to reject the null hypothesis that the market return is exogenous.

```
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
```

(g)The IV/2SLS estimate gives a slightly higher coefficient than the OLS estimate (1.2831 vs. 1.2018). This suggests that the instrument (v_hat2) might be capturing some part of the variation that OLS does not, leading to a higher estimate.

```
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 
```

(h)At the 5% significance level, we fail to reject the null hypothesis, which means that both RANK and POS are valid instrumental variables and can be considered exogenous variables.

```
R squared= 0.003102574
nRsquared= 0.5584634
p-value= 0.45488
```

Q10.24:

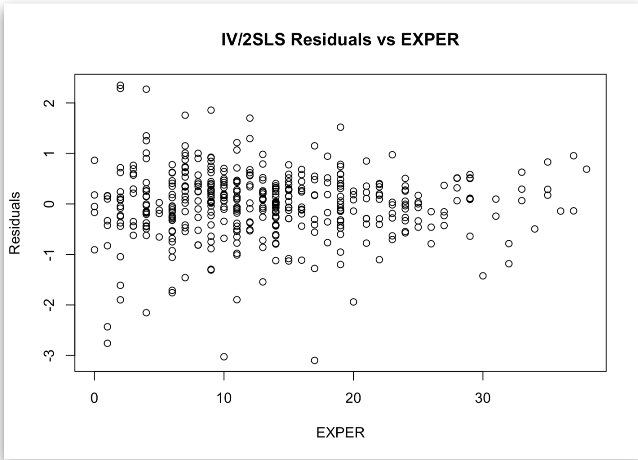
(a)The points appear to be randomly scattered without any visible trend or systematic structure, which suggests that the assumption of homoskedasticity holds.

```
Call:
ivreg(formula = lwage ~ exper + exper2 + educ | exper + exper2 + 
    mothereduc + fathereduc, data = mroz_sub)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.0986 -0.3196  0.0551  0.3689  2.3493 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)   
(Intercept)  0.0481003  0.4003281   0.120  0.90442   
exper        0.0441704  0.0134325   3.288  0.00109 **
exper2      -0.0008990  0.0004017  -2.238  0.02574 * 
educ         0.0613966  0.0314367   1.953  0.05147 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6747 on 424 degrees of freedom
Multiple R-Squared: 0.1357,	Adjusted R-squared: 0.1296 
Wald test: 8.141 on 3 and 424 DF,  p-value: 2.787e-05 
```



(b)We reject the null hypothesis of homoskedasticity.

```
> nR2
[1] 7.438552
> p_value
[1] 0.006384122
```

(c)The 95% Confidence Interval (using Robust SE)=[−0.00393,0.12673]

```
            Estimate Baseline_SE Robust_SE Increased_SE
(Intercept)  0.04810     0.40033   0.42980          Yes
exper        0.04417     0.01343   0.01555          Yes
exper2      -0.00090     0.00040   0.00043          Yes
educ         0.06140     0.03144   0.03334          Yes
```

(d)Bootstrap SE is larger than the baseline SE but smaller than the robust SE.
The 95% Confidence Interval (using Bootstrap SE)=[-0.002,0.1248]

```
> cat("  95% CI for EDUC (bootstrap) = [",
+     round(ci_boot[1],4), ", ", round(ci_boot[2],4), "]\n")
  95% CI for EDUC (bootstrap) = [ -0.002 ,  0.1248 ]
```