Student: Le Thi Phuong Thao – 413707007

Exercise 10.18

a) Answer

```
> print(paste("Percentage of mothers with some college education:", percentage_mothercoll, "%"))
[1] "Percentage of mothers with some college education: 12.1495327102804 %"
> print(paste("Percentage of fathers with some college education:", percentage_fathercoll, "%"))
[1] "Percentage of fathers with some college education: 11.6822429906542 %"
```

b) Answer

```
> correlation_matrix <- cor(cbind(mroz$educ,mroz$MOTHERCOL,mroz$FATHERCOLL))
> print(correlation_matrix)
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.3081925 0.3362576
[2,] 0.3081925 1.0000000 0.3901653
[3,] 0.3362576 0.3901653 1.0000000
```

EDUC vs. MOTHERCOLL (r = 0.31): Children with more years of education tend somewhat to have mothers who also went past high school.

EDUC vs. FATHERCOLL (r = 0.34): A very similar pattern holds for fathers.

MOTHERCOLL vs. FATHERCOLL (r = 0.39): If one parent has some college, there's a slightly higher chance the other does as well.

None of the correlations exceed 0.4, so there's no strong collinearity among these three variables—just a modest positive relationship in each case.

They can be better IV than Mother_Educ and Father_educ if evidence suggests the main jump comes at "some college" (e.g., gains level off after 2 years); a 0/1 indicator will neatly capture that. If we only care about a policy threshold (e.g. whether parents went on for any post-secondary schooling), the dummy is more directly relevant. It often improves degrees of freedom and robustness if the continuous effect is very non-linear

c) Answer

```
> mroz1 <- mroz[mroz$lfp==1,]
> mroz1$MOTHERCOLL <- ifelse(mroz1$mothereduc > 12, 1, 0)
> wage_iv_model <- ivreg(log(wage) ~ educ + exper + I(exper^2) | MOTHERCOLL + exper + I(exper^2), data = mroz1)
> educ_95percent_interval <- confint(wage_iv_model, level = 0.95)["educ",]
> cat("The 95% interval estimate for the coefficient of EDUC is [",round(educ_95percent_interval,4),"]")
The 95% interval estimate for the coefficient of EDUC is [ -0.0012 0.1533 ]
```

d) Answer: F-test statistic for the hypothesis that MOTHERCOLL has no effect on EDUC: 63.21602 > 10, we reject the null hypothesis that the IV is weak

e) 95% CI of the instrumental variables is [0.02752,0.1482], which is narrower than part c

f) F-test statistic for the joint significance of MOTHERCOLL and FATHERCOLL: 57.59666 > 10, we reject the hypothesis that IV is weak

g) We have one endogenous variable but two IV. We need to test the overidentification of IV. If all surplus moment conditions are valid, then NR2 ~ Chi-square(L−B). Sargan-Hansen statistic for the validity of the surplus instrument: 0.2375851 < Critical value 3.841459. We cannot reject H0 that the null of valid over-identifying restrictions → the surplus instrument appears valid

Exercise 10.20

a) Summary for Microsoft stock

```
Call:
lm(formula = msftrf ~ rmrf, 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
rmrf        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
```

Beta = 1.2 and significant at 1%, Microsoft stock is about 20 % more volatile than the market portfolio

b) Answer

```
Call:
lm(formula = rmrf ~ 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
```

Rank can be a potential IV since it is not related to Microsoft stock but related to market beta. F value of 1st stage is 1858, indicating that Rank seems to be a strong IV

c) Answer

```
Call:
lm(formula = msftrf ~ vhat + rmrf, 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
vhat        -0.874599   0.428626  -2.040   0.0428 *
rmrf         1.278318   0.126749  10.085   <2e-16 ***
---
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
```

Hausman test indicates that at 1% level, we cannot reject H0: market return is exogenous

d) Answer

```
> confint(iv1, level = 0.95)
                  2.5 %     97.5 %
(Intercept) -0.008827134 0.01486322
rmrf         1.027421458 1.52921503
> summary(iv1)

Call:
ivreg(formula = msftrf ~ rmrf | 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
rmrf        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 coefficient of beta in IV regression (1.2783) is slightly larger than the original OLS one (1.2018), which is what we would expect. The 95% interval estimate is now [1.0274, 1.5292]

e) Answer

```
Call:
lm(formula = rmrf ~ 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
```

F-value is 951.3, so IV is not weak, RANK remains strongly significant, but POS is significant at the 5% level. If we accept a 5% test, we can conclude the IV are not weak.

f) Answer

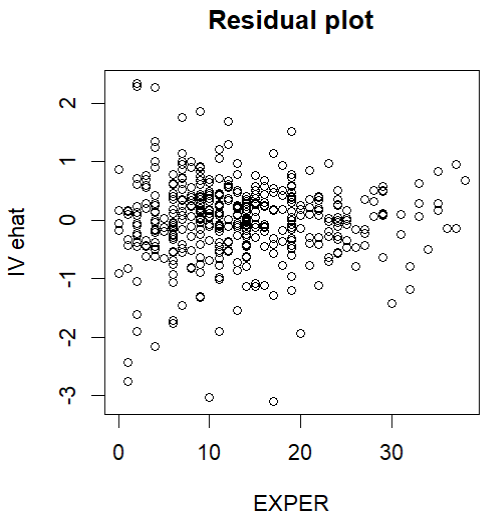
```
Call:
lm(formula = msftrf ~ rmrf + vhat2, data = capm5)

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
rmrf         1.283118   0.126344  10.156   <2e-16 ***
vhat2       -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 hauman test indicates that at 1% level we cannot reject the null hypothesis that the market return is exogenous.

Exercise 10.24

a) Plot ehat IT with expert

When experience is low, the variation in the residuals appears to be larger. This indicates possible heteroskedasticity.



b) Answer

```
> aux_reg <- lm(ehat_2 ~ mroz2$exper)
> n <- length(ehat_2)
> R2 <- summary(aux_reg)$r.squared
> NR2_stat <- n * R2
> NR2_stat <- n * R2
> pchisq(NR2_stat, df = 1, lower.tail = FALSE)
[1] 0.006384122
> p_value <- pchisq(NR2_stat, df = 1, lower.tail = FALSE)
> cat("NR2 test statistic =", NR2_stat, "\n")
NR2 test statistic = 7.438552
> cat("p-value =", p_value, "\n")
p-value = 0.006384122
```

NR2 test statistics = 7.438552

p-value = 0.006384122

It indicates strong evidence of heteroskedasticity in the model

c) 95% CI for the coefficient of 'educ': [-0.0003945456 , 0.1231878]
95% CI for the coefficient of 'educ' with robust SE: [-0.004764123 , 0.1275574]

d)