Q10.18.a

	prop_fathercoll	prop_mothercoll
1	0.1075697	0.1009296

	prop_either_coll	prop_both_coll
1	0.1633466	0.04515272

Beside shows the sample proportion of
father having college degree in the obs.
mother :
Either father or mother :
Both father and mother :

Q10.18.b

	educ	mothercoll	fathercoll
educ	1.00	0.34	0.32
mothercoll	0.34	1.00	0.37
fathercoll	0.32	0.37	1.00

Beside shows the correlation matrix of the three variables

$$Wage = \alpha + \beta\, Educ + u$$

$$Cov(Z, u) = 0$$

$$Cov(Z, Educ) > 0$$

$$Corr(Z, Educ) \uparrow\uparrow\uparrow$$

The endogeneity of Educ on Wage may be caused by unobserved family background, and the continuous Parent education may catch the variation of the unobserved family background, making $Cov(Z, u) \neq 0$. A dummy might reduce the risk of $Cov(Z, u) \neq 0$ since it brings less covariation by filtering out detail information of Parent education.

Q10.18.c

Via 2SLS regression, → With mothercoll as IV the obtained Confident interval of each slope coefficient shows in the table

	2.5 %	97.5 %
(Intercept)	-1.105942034	8.404298e-01
educ	-0.001219763	1.532557e-01
exper	0.017054428	6.963439e-02
I(exper^2)	-0.001658392	-8.385898e-05

95% CI for coef of educ = [-0.0012, 1.5356]

→ Not significant @ 95% level

Q 10.18. d

By F test, mothercoll brings excess explanatory power to Educ so that F stat is significant, rejecting H_0 $\beta_{mothercoll} = 0$

Fit IV2 $Cov(Z, X) > 0$

```
Linear hypothesis test:
mothercoll = 0

Model 1: restricted model
Model 2: educ ~ mothercoll + exper + I(exper^2)

  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
```

Q 10.18. e

Via 2SLS reg, using fathercoll & mothercoll as IVs, the CI for all coefficient shows beside.

	2.5 %	97.5 %
(Intercept)	-1.04782153	4.896578e-01
educ	0.02751845	1.481769e-01
exper	0.01661839	6.873386e-02
I(exper^2)	-0.00162779	-6.940599e-05

95% CI for $\beta_{Educ} = [0.028, 1.482]$

Significant @ 95% level

Q 10.18. f

By F test, we reject the null that both mothercoll and fathercoll are not correlated to educ.

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

Model 1: restricted model
Model 2: educ ~ mothercoll + fathercoll + exper + I(exper^2)

  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
```

Null: $Cov(mothercoll, e) = 0$
$Cov(fathercoll, e) = 0$

Q 10.18. g

Through Sargn test,

$q = nR^2 \sim \chi^2_{L-B}$

```
> cat(chicrit , nr2)
3.841459 0.6513957
```

$q^{*} < \chi^2_{crit}$

$\rightarrow$ Doesn't reject H_0 that all IV is valid.

Q 10.20. a

The market beta of msft > 1. meaning it bears a high market risk and is identified as risky.

```
Call:
lm(formula = msft ~ mkt, data = rp)

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         1.201840   0.122152   9.839   <2e-16
               
(Intercept)    
mkt         ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Q. 10.20. b

By model diagnostic test,

IV_1 : At least 1 IV is strong

IV_2 : mkt is endogenous

IV_3 : The model isn't "Over-identified" so the sargan test is not valid.

```
Call:
ivreg(formula = msft ~ mkt | rank, data = rp_sort)

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         1.278318   0.128011   9.986   <2e-16
               
(Intercept)    
mkt         ***

Diagnostic tests:
                 df1 df2 statistic p-value    
Weak instruments   1 178  1857.587  <2e-16 ***
Wu-Hausman         1 177     4.164  0.0428 *  
Sargan             0  NA        NA      NA    
```

In the first stage regression, the R^2 is 0.912, indicating a very strong explanation of Rank on mkt. Plus, it passes the weak-instrument test. We can conclude Rank is strong IV.

```
Call:
lm(formula = mkt ~ rank, data = rp_sort)

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

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

Residual standard error: 0.01467 on 178 degrees of fr
eedom
Multiple R-squared:  0.9126,	Adjusted R-squared:  
0.9121 
```

Q 10.20.c

By Hausman test,

$$y = \beta_1 + \beta_2 X + e$$

$$X = \gamma_1 + \theta_1 z + v$$

$$\underbrace{Cov(X, e) \neq 0}_{\text{endogenous}} \to Cov(v, e) \neq 0 \to \underbrace{\gamma \neq 0}_{\text{endogenous}}$$

$$y = \beta_1 + \beta_2 X + \gamma \underbrace{\hat{v} + u}_{e}$$

H₀: $\gamma = 0 \to$ X is exogenous

H₁: $\gamma \neq 0 \to$ X is endogenous

```
                  0.5 %      99.5 %
(Intercept) -0.0125636 0.01859968
mkt          0.9482782 1.60835833
v_hat       -1.9906947 0.24149698
```

→ 99% CI for $\gamma = [-1.99, 0.241]$

Showing mkt is exogenous under $\alpha = 1\%$

Q.10.20.d

When there's measurement error in mkt, ,

$Mkt = X + e$

Unobserved real exogenous market portfolio

$$R_i = \alpha + \beta X + u$$

Mkt

measurement error

$$R_i = \alpha + \beta(X + e) + u$$

$$= \alpha + \beta X + \beta e + u$$

$$\beta = \frac{Cov(X + e, \alpha + \beta X + \beta e + u)}{Var(X + e)}$$

$$= \boxed{\beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2}} < \boxed{\beta}$$

$\hat{\beta}_{LM} \xrightarrow{P} \beta \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2}$ $\qquad$ $\hat{\beta}_{IV} \xrightarrow{P} \beta$

LM with measurement error

```
Call:
lm(formula = msft ~ mkt, data = rp)

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         1.201840   0.122152   9.839   <2e-16

(Intercept)
mkt         ***
```

2SLS

```
Call:
ivreg(formula = msft ~ mkt | rank, data = rp_sort)

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         1.278318   0.128011   9.986   <2e-16

(Intercept)
mkt         ***
```

1.2 < 1.27 → agree with expectation !!

Q.10.20.e

In first stage mkt on rank and pos, the p-value of joint F test is significant under each level of α and $R^2 = 0.915$, showing the two IV are strong.

```
Call:
lm(formula = mkt ~ rank + pos, data = rp_sort)

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

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

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

Q10.18.f,g

According to Hausman test, @ 1% level, we cannot reject the null that mkt is exogenous. For Sargan test @ 1% level, the IV's are valid

```
Call:
ivreg(formula = msft ~ mkt | rank + pos, data = rp_so
rt)

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         1.283118   0.127866  10.035   <2e-16
               
(Intercept)    
mkt         ***

Diagnostic tests:
                 df1 df2 statistic p-value    
Weak instruments   2 177   951.262  <2e-16 ***
Wu-Hausman         1 177     4.862  0.0287 *  
Sargan             1  NA     0.558  0.4549    
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

H_0: Mkt is exogenous

H_0: IV's are valid

Q10.18.g

by (d)

$\hat{\beta}_{LM} < \hat{\beta}_{IV}$

$1.2 < 1.28$

agree the expectation

$\hat{\beta}_{IV}$

```
Call:
ivreg(formula = msft ~ mkt | rank + pos, data = rp_sc
rt)

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         1.283118   0.127866  10.035   <2e-16
               
(Intercept)    
mkt         ***
```

$\hat{\beta}_{LM}$

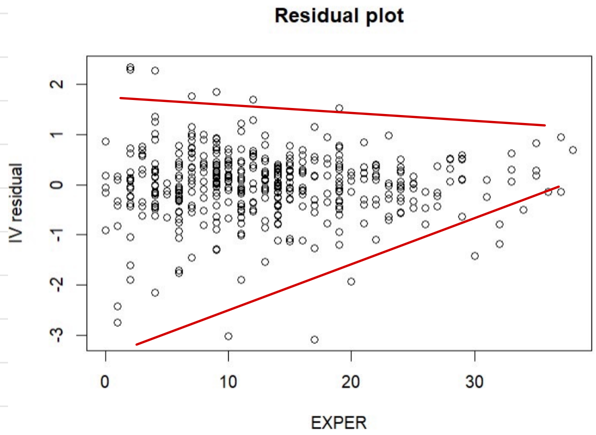
```
Call:
lm(formula = msft ~ mkt, data = rp)

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         1.201840   0.122152   9.839   <2e-16
               
(Intercept)    
mkt         ***
```

Q10.24.a

Errors has a pattern of heteroskedasticity with Exper



Q10.24.b

By BP test, we reject the null that the error is homoskedastic with exper under any level of α.

```
> bptest(modiv,
+        varformula = ~ exper,
+        data = data)

	studentized Breusch-Pagan test

data:  modiv
BP = 7.7985, df = 1, p-value = 0.005229
```

Q10.24.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 ]
```

The gap becomes wider for robust SE

→ Robust SE > Normal SE

Q10.24.d

```
Bootstrap 95% CI for the coefficient of 'educ':[ -0.01016399 , 0.129295 ]
```

The boostraped CI become even wider the the other methods.