10.18

(a)

```
> mean(data$MOTHERCOLL)
[1] 0.1214953
> mean(data$FATHERCOLL)
[1] 0.1168224
```

(b)

```
> cor(data$educ, data$MOTHERCOLL, use = "complete.obs")
[1] 0.3594705
> cor(data$educ, data$FATHERCOLL, use = "complete.obs")
[1] 0.3984962
```

EDUC is positively correlated with the parents' dummy variables.

(c)

```
> confint(IV1, "educ", level = 0.95)
              2.5 %     97.5 %
educ -0.001219763 0.1532557
```

(d) The first-stage equation for one IV

EDUC = 12.079094 + 0.05623 EXPER – 0.001956 EXPER^2 + 2.517068 MOTHERCOLL + v

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

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 ***
MOTHERCOLL   2.517068   0.315713   7.973 1.46e-14 ***
exper        0.056230   0.042101   1.336    0.182
I(exper^2)  -0.001956   0.001256  -1.557    0.120
---
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
```

$H_0: \beta_{MOTHERCOLL} = 0$

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

```
Anova Table (Type II tests)

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

The F-value for MOTHERCOLL is 63.5631 >10, so we fail to reject H0, which means that MOTHERCOLL has a significant effect on EDUC. MOTHERCOLL is a strong IV.

(e)

```
> confint(IV2, "educ", level = 0.95)
             2.5 %     97.5 %
educ 0.02751845 0.1481769
```

It's narrower from part (c).

(f)

```
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
```

The F-test statistic of the joint significance of MOTHERCOLL and FATHERCOLL is 56.963>10, so we reject H0 which means the IV are weak.

(g)

```
Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | MOTHERCOLL +
    FATHERCOLL + exper + I(exper^2), data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-3.07797 -0.32128  0.03418  0.37648  2.36183

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.2790819  0.3922213  -0.712  0.47714
educ         0.0878477  0.0307808   2.854  0.00453 **
exper        0.0426761  0.0132950   3.210  0.00143 **
I(exper^2)  -0.0008486  0.0003976  -2.135  0.03337 *

Diagnostic tests:
                 df1 df2 statistic p-value
Weak instruments   2 423    56.963  <2e-16 ***
Wu-Hausman         1 423     0.519   0.472
Sargan             1  NA     0.238   0.626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6679 on 424 degrees of freedom
Multiple R-Squared: 0.153,      Adjusted R-squared: 0.147
Wald test: 9.724 on 3 and 424 DF,  p-value: 3.224e-06
```

The p-value is really small, which means that the IV do not have any issues in the regression model and there are no overidentifying restrictions.

10.20

(a) Microsoft stock is risky.

```
Call:
lm(formula = msft_rp ~ mkt_rp, data = combined_data)

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_rp      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)

```
Call:
lm(formula = mkt_rp ~ RANK, data = combined_data)

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
```

Because F statistic is larger than 10, it means RANK is a strong IV.

(c)

```
Call:
lm(formula = msft_rp ~ mkt_rp + v_hat, data = combined_data)

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_rp       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

>
> p_value <- coef(summary(mod_aux))["v_hat", "Pr(>|t|)"]
> if(p_value < 0.01){
+   cat("We reject H0, it means market risk premium is endogeneity.\n")
+ } else{
+   cat("We fail to reject H0, it means market risk premium might be exogeneity.\n")
+ }
We fail to reject H0, it means market risk premium might be exogeneity.
```

(d)

	Model	Beta	SE	R2
1	OLS	1.201840	0.1221516	0.3522665
2	2SLS	1.278318	0.1280109	0.3508400

(e)

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

Model 1: restricted model
Model 2: mkt_rp ~ 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 ***
```

Because F statistic is larger than 10, it means RANK is a strong IV.

```
Call:
lm(formula = mkt_rp ~ RANK + POS, data = combined_data)

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)

```
Call:
lm(formula = msft_rp ~ mkt_rp + v_hat, data = combined_data)

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_rp       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
```

Because the p-value of v_hat > 0.01, we fail to reject H0, it means that the market return is exogenous at the 1% level of significance.

(g)

```
Call:
lm(formula = msft_rp ~ mkt_rp, data = combined_data)

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_rp      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
```

The IV/2SLS estimate gives a slightly higher coefficient than the OLS estimate(1.2018 & 1.2831). This suggests that instrument (v_hat2) might be capturing some part if variation the OLS does nit, leading to a higher estimate.

```
Call:
ivreg(formula = msft_rp ~ mkt_rp | RANK + POS, data = combined_data)

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_rp      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
```

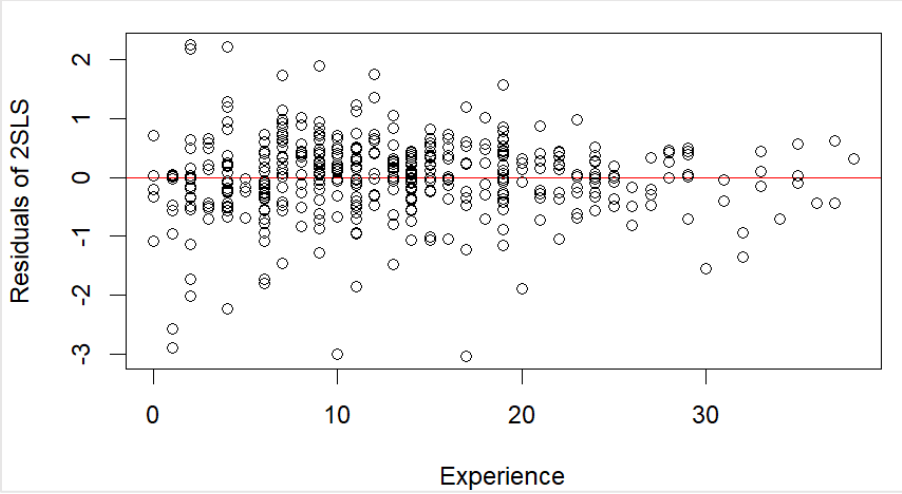
(h)

```
> sargan_stat
[1] 0.5584634
> p_value_sargan
[1] 0.45488
```

Because the p-value of Sargan test is > 0.05, we fail to reject H0, it means that both RANK and POS are valid IV and can be considered exogenous variables.

10.24

(a)The points seems to be randomly scattered, which suggests the assumption of homoskedasticity holds.



(b)

```
> nr2_test
[1] 8.40036
> p_value
[1] 0.003751468
```

Because the p-value < 0.05, we reject H0, it means it the heteroskedasticity exists.

(c)

```
            Estimate Baseline_SE Robust_SE Increased_SE
(Intercept)  0.14784     0.40221   0.42922          Yes
educ         0.06639     0.03125   0.03358          Yes
exper        0.01549     0.00406   0.00414          Yes
> cat("Conclusion:\nRobust SEs are larger, indicating heteroskedasticity.\n")
Conclusion:
Robust SEs are larger, indicating heteroskedasticity.
> cat(sprintf("95%% Robust CI for EDUC: [%.4f, %.4f]\n", ci[1], ci[2]))
95% Robust CI for EDUC: [0.0006, 0.1322]
```

(d)

```
95% CI for EDUC (bootstrap) = [ -0.002 ,  0.1248 ]
```