Full name: Nguyen Nhut Vu Truong

Student ID: 413707008

Course: Financial Econometrics

HW0428

Question 18

a.

```
> # a
> # Create dummy variables for mother's and father's college education
> mroz <- mroz %>%
+   mutate(
+     MOTHERCOLL = ifelse(mothereduc > 12, 1, 0),
+     FATHERCOLL = ifelse(fathereduc > 12, 1, 0)
+   )
>
> # Calculate the percentage of parents with some college education
> percentage_mother = mean(mroz$MOTHERCOLL) * 100
> percentage_father = mean(mroz$FATHERCOLL) * 100
>
> cat("Percentage of mothers with some college education:", percentage_mother, "%\n")
Percentage of mothers with some college education: 12.14953 %
> cat("Percentage of fathers with some college education:", percentage_father, "%\n")
Percentage of fathers with some college education: 11.68224 %
```

b.

MOTHERCOLL and FATHERCOLL are binary variables, reducing potential measurement error compared to continuous variables (MOTHEREDUC, FATHEREDUC). It is easier to explain these two variables: Whether having a high degree of education.

```
> # b
> # Correlations between EDUC, MOTHERCOLL, and FATHERCOLL
> correlations <- cor(mroz %>% select(educ, MOTHERCOLL, FATHERCOLL), use = "complete.obs")
> print(correlations)
                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.

```
> # Part (c): IV regression using MOTHERCOLL as the instrument for educ
> iv_c <- ivreg(log(wage) ~ educ + exper + I(exper^2) | MOTHERCOLL + exper + I(exper^2), data = mroz)
>
> # Summary with robust standard errors
> summary(iv_c, vcov = sandwich)

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

Residuals:
     Min       1Q   Median       3Q      Max
-3.08719 -0.32444  0.04147  0.36634  2.35621

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1327561  0.5607243  -0.237   0.8130
educ         0.0760180  0.0437323   1.738   0.0829 .
exper        0.0433444  0.0152145   2.849   0.0046 **
I(exper^2)  -0.0008711  0.0004175  -2.086   0.0375 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6703 on 424 degrees of freedom
Multiple R-Squared: 0.147,	Adjusted R-squared: 0.1409
Wald test: 6.149 on 3 and 424 DF,  p-value: 0.0004237
```

```
>
> # 95% confidence interval for educ coefficient
> confint(iv_c, 'educ', level = 0.95)
           2.5 %    97.5 %
educ -0.001219763 0.1532557
```

d.

Reject the null hypothesis that the coefficient on MOTHERCOLL is zero. MOTHERCOLL is a **strong instrument** for educ based on both tests.

```
> # H0: coefficient on MOTHERCOLL = 0
> linearHypothesis(first_stage, "MOTHERCOLL = 0", vcov = vcovHC(first_stage, type = "HC1"))

Linear hypothesis test:
MOTHERCOLL = 0

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

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)
1    425
2    424  1 76.392 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> linearHypothesis(first_stage, "MOTHERCOLL = 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
```

e. The new CI is narrower, then adding FATHERCOLL has improved precision (likely due to stronger instrumentation).

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

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  0.4302159  -0.649  0.51688
educ         0.0878477  0.0337910   2.600  0.00966 **
exper        0.0426761  0.0154095   2.769  0.00586 **
I(exper^2)  -0.0008486  0.0004255  -1.995  0.04673 *
---
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: 7.693 on 3 and 424 DF,  p-value: 5.136e-05
```

```
> # 95% confidence interval for educ coefficient
> confint(iv_c, 'educ', level = 0.95)
            2.5 %    97.5 %
educ -0.001219763 0.1532557
> # 95% confidence interval for the coefficient on educ
> confint(iv_e, 'educ', level = 0.95)
          2.5 %    97.5 %
educ 0.02751845 0.1481769
```

f. **Both results are statistically significant** at the 1% level (***), meaning we reject the null hypothesis that MOTHERCOLL and FATHERCOLL jointly have no effect on educ. The **F-statistics are well above 10** (a common rule-of-thumb threshold), indicating **strong instruments**.

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

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 ***
MOTHERCOLL   1.749947   0.322347   5.429 9.58e-08 ***
FATHERCOLL   2.186612   0.329917   6.628 1.04e-10 ***
exper        0.049149   0.040133   1.225    0.221
I(exper^2)  -0.001449   0.001199  -1.209    0.227
---
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 ~ MOTHERCOLL + FATHERCOLL + exper + I(exper^2)

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)
1    425
2    423  2 89.329 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> linearHypothesis(first_stage_e,
+                  c("MOTHERCOLL = 0", "FATHERCOLL = 0"))

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
```

g. The Sargan test yields a p-value of 0.626, indicating that we fail to reject the null hypothesis of instrument validity, and thus both MOTHERCOLL and FATHERCOLL appear to be valid instruments for EDUC in the wage equation.

```
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
```

Question 20

a.

beta = 1.201840 => risky relative to the market portfolio

```
Call:
lm(formula = rp_msft ~ rp_mkt, 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    
rp_mkt      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. To assess IV conditions:

- **IV1-IV3** refer to: relevance, exogeneity, and exclusion.
- RANK is deterministic from rp_mkt, so it may be **relevant** (IV1), but its validity as **exogenous** (IV2) is debatable.

R2 = 0.9126, Adjusted R2 = 0.9121

Reject the null hypothesis that the coefficient on RANK is zero (F-value = 1857.6). RANK is a **strong instrument** for educ based on both tests.

A rule of thumb is that an F-statistic > 10 suggests a strong instrument (Staiger & Stock, 1997).

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

c. Because **p = 0.0428 > 0.01**, we conclude that **market return is exogenous** in the CAPM for Microsoft.

```
Call:
lm(formula = rp_msft ~ rp_mkt + 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    
rp_mkt       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 estimate of β (1.2783) is slightly higher than the OLS estimate (1.2018), both highly significant, suggesting little endogeneity in the market return and indicating that OLS provides reliable results consistent with CAPM expectations.

Method	**Estimate of β (rp_mkt)**	**Std. Error**	**t-value**	**Significance**
OLS	1.2018	0.1222	9.84	***
IV	1.2783	0.1280	9.99	***

e. R2 = 0.9149, Adjusted R2 = 0.9139

Since the **F-statistic > 10**, the instruments are considered **jointly strong**.

```
Call:
lm(formula = rp_mkt ~ 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: rp_mkt ~ 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. The null hypothesis H_0: $\delta = 0$ (market return is exogenous) is tested at the 1% significance level. Since **p-value = 0.0287 > 0.01**, **fail to reject H_0** → Market return is **exogenous.**

```
Call:
lm(formula = rp_msft ~ rp_mkt + v_hat_e, 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    
rp_mkt       1.283118   0.126344  10.156   <2e-16 ***
v_hat_e     -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 estimate of β (1.2831) using both RANK and POS is slightly higher than the OLS estimate (1.2018), consistent with expectations under potential measurement error in the market return, and further supports the reliability of OLS given earlier evidence of exogeneity.

Method	**β Estimate (rp_mkt)**	**Std. Error**	**t-value**	**Conclusion**
OLS	1.2018	0.1222	9.84	Significant
IV (RANK + POS)	**1.2831**	0.1279	10.04	Significant

```
Call:
ivreg(formula = rp_msft ~ rp_mkt | 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    
rp_mkt      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. We first compute the residuals from the IV/2SLS model in part (g), then regress them on all instruments (RANK and POS), and finally use the NR^2 statistic from this regression to perform the Sargan test for instrument validity.

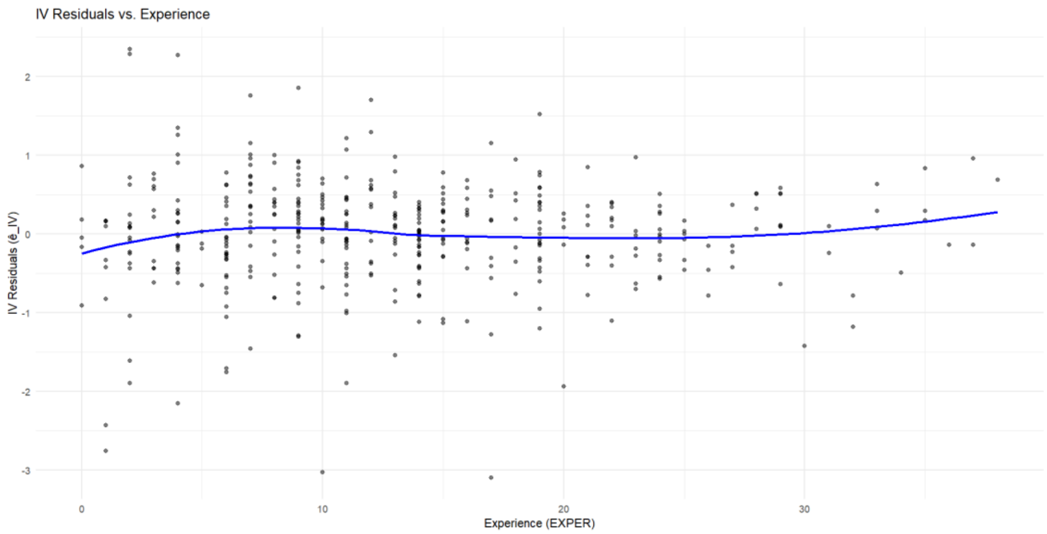
The Sargan test yields a statistic of 0.5585 with a p-value of 0.4549, so we fail to reject the null hypothesis at the 5% level and conclude that the surplus instrument POS is valid and the instruments as a whole are exogenous.

```
> cat("Sargan statistic:", sargan_stat, "\n")
Sargan statistic: 0.5584634
> cat("p-value:", p_value, "\n")
p-value: 0.45488
> summary(iv_model_g)

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
---
```

Question 24

a. The residual plot shows no clear pattern of increasing or decreasing spread with experience, indicating that the residuals are **generally consistent with homoskedasticity**.



b.

Null hypothesis H0: Homoskedasticity (constant variance)

Since the Breusch-Pagan test yields a statistic of 7.44 with a p-value of 0.0064 (< 0.01), we **reject the null hypothesis** of homoskedasticity and conclude that there is **evidence of heteroskedasticity** in the IV residuals.

```
Call:
lm(formula = e_iv_sq ~ exper, data = mroz)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.6740 -0.4341 -0.2685 -0.0168  9.2188 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.676563   0.096573   7.006 9.65e-12 ***
exper       -0.017303   0.006303  -2.745  0.00631 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.049 on 426 degrees of freedom
Multiple R-squared:  0.01738,	Adjusted R-squared:  0.01507 
F-statistic: 7.535 on 1 and 426 DF,  p-value: 0.006308

> N <- nrow(mroz)
> R2_bp <- summary(bp_model)$r.squared
> bp_stat <- N * R2_bp
> p_val <- 1 - pchisq(bp_stat, df = 1)
> 
> cat("Breusch-Pagan test statistic (NR²):", bp_stat, "\n")
Breusch-Pagan test statistic (NR²): 7.438552 
> cat("p-value:", p_val, "\n")
p-value: 0.006384122
```

c. The **robust standard error is slightly larger**, which is expected under heteroskedasticity (0.0333 > 0.0314). The confidence interval for educ is slightly **wider** using heteroskedasticity-robust standard errors (−0.0041 to 0.1269) than with baseline standard errors (−0.0004 to 0.1232), reflecting the adjustment for potential heteroskedasticity in the IV model.

```
> coefci(iv_model, parm = "educ", level = 0.95,
+                    vcov. = vcovHC(iv_model, type = "HC1"))   # 95% CI using Robust SE
            2.5 %    97.5 %
educ -0.004132858 0.1269261
> coefci(iv_model, parm = "educ", level = 0.95)  # 95% CI using Baseline SE
             2.5 %    97.5 %
educ -0.0003945456 0.1231878

> # Robust standard errors
> coeftest(iv_model, vcov = vcovHC(iv_model, type = "HC1"))

t test of coefficients:

               Estimate  Std. Error t value Pr(>|t|)   
(Intercept)  0.04810030  0.42979772  0.1119 0.910945   
educ         0.06139663  0.03333859  1.8416 0.066231 . 
exper        0.04417039  0.01554638  2.8412 0.004711 **
I(exper^2)  -0.00089897  0.00043008 -2.0902 0.037193 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> summary(iv_model)

Call:
ivreg(formula = log(wage) ~ educ + exper + I(exper^2) | mothereduc +
    fathereduc + exper + I(exper^2), data = mroz)

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
educ         0.0613966  0.0314367   1.953  0.05147 .
exper        0.0441704  0.0134325   3.288  0.00109 **
I(exper^2)  -0.0008990  0.0004017  -2.238  0.02574 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> confint(iv_model, "educ", level = 0.95, vcov. = vcovHC(iv_model, type = "HC1"))
             2.5 %    97.5 %
educ -0.0002181633 0.1230114
```

d.

The bootstrap standard error for educ is **0.0323**, which is slightly **larger** than the baseline SE (0.0314) but slightly **smaller** than the robust SE (0.0333); using the bootstrap SE and the original model's coefficient (0.0614), the **95% confidence interval** is approximately **[−0.0020, 0.1248]**.

The bootstrap standard errors are slightly larger than the baseline SEs and comparable to the robust SEs for all coefficients, with educ having a bootstrap SE of 0.0323 versus 0.0314 (baseline) and 0.0333 (robust), indicating consistent but slightly more conservative inference under heteroskedasticity.

```
> # Display the comparison table
> print(se_table)
            Coefficient Baseline_SE Robust_SE Bootstrap_SE
(Intercept) (Intercept)      0.4003    0.4298       0.4379
educ               educ      0.0314    0.0333       0.0323
exper             exper      0.0134    0.0156       0.0158
I(exper^2)   I(exper^2)      0.0004    0.0004       0.0004

> # Display
> cat("Point estimate for EDUC (original model):", beta_hat, "\n")
Point estimate for EDUC (original model): 0.06139663
> cat("Bootstrap SE:", boot_se, "\n")
Bootstrap SE: 0.03234547
> cat("95% CI using bootstrap SE:", ci_boot, "\n")
95% CI using bootstrap SE: -0.002000496 0.1247938
```