10.18 Consider the data file *mroz* on working wives. Use the 428 observations on married women who participate in the labor force. In this exercise, we examine the effectiveness of a parent's college education as an instrumental variable.

a. Create two new variables. *MOTHERCOLL* is a dummy variable equaling one if *MOTHEREDUC* > 12, zero otherwise. Similarly, *FATHERCOLL* equals one if *FATHEREDUC* > 12 and zero otherwise. What percentage of parents have some college education in this sample?

b. Find the correlations between *EDUC*, *MOTHERCOLL*, and *FATHERCOLL*. Are the magnitudes of these correlations important? Can you make a logical argument why *MOTHERCOLL* and *FATHERCOLL* might be better instruments than *MOTHEREDUC* and *FATHEREDUC*?

c. Estimate the wage equation in Example 10.5 using *MOTHERCOLL* as the instrumental variable. What is the 95% interval estimate for the coefficient of *EDUC*?

d. For the problem in part (c), estimate the first-stage equation. What is the value of the *F*-test statistic for the hypothesis that *MOTHERCOLL* has no effect on *EDUC*? Is *MOTHERCOLL* a strong instrument?

e. Estimate the wage equation in Example 10.5 using *MOTHERCOLL* and *FATHERCOLL* as the instrumental variables. What is the 95% interval estimate for the coefficient of *EDUC*? Is it narrower or wider than the one in part (c)?

f. For the problem in part (e), estimate the first-stage equation. Test the joint significance of *MOTHERCOLL* and *FATHERCOLL*. Do these instruments seem adequately strong?

g. For the IV estimation in part (e), test the validity of the surplus instrument. What do you conclude?

a.

```
父母中至少有一方受過大學教育的百分比： 16.33 %
```

b.

```
> # 輸出相關係數
> cat("educ 與 mothercoll 的相關係數：", round(cor_educ_mothercoll, 3), "\n")
educ 與 mothercoll 的相關係數： 0.337
> cat("educ 與 fathercoll 的相關係數：", round(cor_educ_fathercoll, 3), "\n")
educ 與 fathercoll 的相關係數： 0.319
> cat("mothercoll 與 fathercoll 的相關係數：", round(cor_mothercoll_fathercoll, 3), "\n")
mothercoll 與 fathercoll 的相關係數： 0.367
```

C.

```
> # 輸出結果
> cat("educ 的係數估計：", round(educ_coef, 3), "\n")
educ 的係數估計： 0.482
> cat("educ 係數的 95% 置信區間：(", round(ci_lower, 3)
educ 係數的 95% 置信區間：( 0.196 ,  0.768 )
```

d.

```
> # 輸出 F 統計量
> cat("假設 mothercoll 對 educ 無影響的 F 檢驗統計量：", rou
假設 mothercoll 對 educ 無影響的 F 檢驗統計量： 96.229

F 統計量 > 10，mothercoll 可能是強工具變量。
```

e.

```
> cat("educ 的係數估計：", round(educ_coef, 3
educ 的係數估計： 0.46
> cat("educ 係數的 95% 置信區間：(", round(ci_
educ 係數的 95% 置信區間：( 0.217 ,  0.702 )
```

f.

```
> cat("假設 mothercoll 對 educ 無影響的 F 檢驗統計量：",
假設 mothercoll 對 educ 無影響的 F 檢驗統計量： 49.683
>
> # 判斷是否為強工具變量
> if (f_statistic > 10) {
+   cat("F 統計量 > 10，mothercoll 可能是強工具變量。\n"
+ } else {
+   cat("F 統計量 <= 10，mothercoll 可能是弱工具變量。\r
+ }
F 統計量 > 10，mothercoll 可能是強工具變量。
>
```

10.20 The CAPM [see Exercises 10.14 and 2.16] says that the risk premium on security j is related to the risk premium on the market portfolio. That is

$$r_j - r_f = \alpha_j + \beta_j(r_m - r_f)$$

where r_j and r_f are the returns to security j and the risk-free rate, respectively, r_m is the return on the market portfolio, and β_j is the jth security's "*beta*" value. We measure the market portfolio using the Standard & Poor's value weighted index, and the risk-free rate by the 30-day LIBOR monthly rate of return. As noted in Exercise 10.14, if the market return is measured with error, then we face an errors-in-variables, or measurement error, problem.

a. Use the observations on Microsoft in the data file *capm5* to estimate the CAPM model using OLS. How would you classify the Microsoft stock over this period? Risky or relatively safe, relative to the market portfolio?

b. It has been suggested that it is possible to construct an IV by ranking the values of the explanatory variable and using the rank as the IV, that is, we sort $(r_m - r_f)$ from smallest to largest, and assign the values $RANK = 1, 2, \ldots, 180$. Does this variable potentially satisfy the conditions IV1–IV3? Create *RANK* and obtain the first-stage regression results. Is the coefficient of *RANK* very significant? What is the R^2 of the first-stage regression? Can *RANK* be regarded as a strong IV?

c. Compute the first-stage residuals, $\hat{v}$, and add them to the CAPM model. Estimate the resulting augmented equation by OLS and test the significance of $\hat{v}$ at the 1% level of significance. Can we conclude that the market return is exogenous?

d. Use *RANK* as an IV and estimate the CAPM model by IV/2SLS. Compare this IV estimate to the OLS estimate in part (a). Does the IV estimate agree with your expectations?

e. Create a new variable $POS = 1$ if the market return $(r_m - r_f)$ is positive, and zero otherwise. Obtain the first-stage regression results using both *RANK* and *POS* as instrumental variables. Test the joint significance of the IV. Can we conclude that we have adequately strong IV? What is the R^2 of the first-stage regression?

f. Carry out the Hausman test for endogeneity using the residuals from the first-stage equation in (e). Can we conclude that the market return is exogenous at the 1% level of significance?

g. Obtain the IV/2SLS estimates of the CAPM model using *RANK* and *POS* as instrumental variables. Compare this IV estimate to the OLS estimate in part (a). Does the IV estimate agree with your expectations?

h. Obtain the IV/2SLS residuals from part (g) and use them (not an automatic command) to carry out a Sargan test for the validity of the surplus IV at the 5% level of significance.

a.

```
> # 提取 beta 值
> beta <- coef(model)["mkt_excess"]
> cat("Microsoft 股票的 beta 值：", round(beta, 3), "\n")
Microsoft 股票的 beta 值： 1.202
>
> # 判斷風險類別
> if (beta > 1) {
+   cat("Microsoft 股票相對於市場組合風險較高。\n")
+ } else if (beta < 1) {
+   cat("Microsoft 股票相對於市場組合風險較低（相對安全）。\n")
+ } else {
+   cat("Microsoft 股票相對於市場組合風險相當。\n")
+ }
Microsoft 股票相對於市場組合風險較高。
```

b.

```
> # 提取 R^2 值
> r_squared <- summary_first_stage$r.squared
> cat("第一階段回歸的 R^2 值：", round(r_squared, 3), "\n")
第一階段回歸的 R^2 值： 0.913
>
> # 判斷工具變量強度（基於 R^2，輔以 F 統計量）
> f_statistic <- summary_first_stage$fstatistic[1]  # 提取 F 統計量
> cat("F 統計量：", round(f_statistic, 3), "\n")
F 統計量： 1857.587
> if (f_statistic > 10) {
+   cat("F 統計量 > 10，RANK 可能是強工具變量。\n")
+ } else {
+   cat("F 統計量 <= 10，RANK 可能是弱工具變量。\n")
+ }
F 統計量 > 10，RANK 可能是強工具變量。
```

c.

```
> # 檢驗 v_hat 的係數顯著性
> v_hat_pvalue <- coef(summary_augmented)["v_hat", "Pr(>|t|)"]
> cat("v_hat 係數的 p 值：", round(v_hat_pvalue, 4), "\n")
v_hat 係數的 p 值： 0.0428
>
> # 在 1% 顯著性水平下檢驗
> if (v_hat_pvalue < 0.01) {
+   cat("在 1% 顯著性水平下，v_hat 顯著，市場回報 (mkt_excess) 內生。\n")
+ } else {
+   cat("在 1% 顯著性水平下，v_hat 不顯著，無法拒絕市場回報 (mkt_excess) 外生的假設。\n")
+ }
在 1% 顯著性水平下，v_hat 不顯著，無法拒絕市場回報 (mkt_excess) 外生的假設。
```

d.

```
> cat("OLS beta 與 IV beta 比較：\n")
OLS beta 與 IV beta 比較：
> if (iv_beta > ols_beta) {
+   cat("IV beta > OLS beta，可能表明 OLS 因內生性向下偏誤。\n")
+ } else if (iv_beta < ols_beta) {
+   cat("IV beta < OLS beta，可能表明其他因素影響估計。\n")
+ } else {
+   cat("IV beta ≈ OLS beta，內生性影響可能不大。\n")
+ }
IV beta > OLS beta，可能表明 OLS 因內生性向下偏誤。
>
> # 判斷 IV 估計是否符合期望
> cat("期望判斷：由於 r_m - r_f 可能內生，IV 估計應修正 OLS 偏誤。如果 IV beta 顯著大於 OLS beta，符合內
生性假設。\n")
期望判斷：由於 r_m - r_f 可能內生，IV 估計應修正 OLS 偏誤。如果 IV beta 顯著大於 OLS beta，符合內生性假
設。
```

e.

```
> cat("第一階段回歸的 R^2 值：", round(r_squared, 3), "\n")
第一階段回歸的 R^2 值： 0.915
>
> # 提取 F 統計量
> f_statistic <- summary_first_stage$fstatistic[1]
> cat("F 統計量：", round(f_statistic, 3), "\n")
F 統計量： 951.262
>
> # 判斷工具變量強度
> if (f_statistic > 10) {
+   cat("F 統計量 > 10，RANK 和 POS 可能是強工具變量。\n")
+ } else {
+   cat("F 統計量 <= 10，RANK 和 POS 可能是弱工具變量。\n")
+ }
F 統計量 > 10，RANK 和 POS 可能是強工具變量。
```

f.

```
> # 檢驗 v_hat 的係數顯著性
> v_hat_pvalue <- coef(summary_augmented)["v_hat", "Pr(>|t|)"]
> cat("v_hat 係數的 p 值：", round(v_hat_pvalue, 4), "\n")
v_hat 係數的 p 值： 0.0287
>
> # 在 1% 顯著性水平下檢驗
> if (v_hat_pvalue < 0.01) {
+   cat("在 1% 顯著性水平下，v_hat 顯著，市場回報 (mkt_excess) 內生。\n")
+ } else {
+   cat("在 1% 顯著性水平下，v_hat 不顯著，無法拒絕市場回報 (mkt_excess) 外生的假設。\n")
+ }
在 1% 顯著性水平下，v_hat 不顯著，無法拒絕市場回報 (mkt_excess) 外生的假設。
```

g.

```
OLS beta 與 IV beta 比較：
> if (iv_beta > ols_beta) {
+   cat("IV beta > OLS beta，可能表明 OLS 因內生性向下偏誤。\n")
+ } else if (iv_beta < ols_beta) {
+   cat("IV beta < OLS beta，可能表明其他因素影響估計。\n")
+ } else {
+   cat("IV beta ≈ OLS beta，內生性影響可能不大。\n")
+ }
IV beta > OLS beta，可能表明 OLS 因內生性向下偏誤。
>
> # 判斷 IV 估計是否符合期望
> cat("期望判斷：由於 r_m - r_f 可能內生（測量誤差），IV 估計應修正 OLS 偏誤。如果 IV beta 顯著大於 OLS b
eta，符合內生性假設。\n")
期望判斷：由於 r_m - r_f 可能內生（測量誤差），IV 估計應修正 OLS 偏誤。如果 IV beta 顯著大於 OLS beta，符
合內生性假設。
```

h.

```
> # 輸出結果
> cat("Sargan 檢驗結果：\n")
Sargan 檢驗結果：
> cat("Sargan 統計量：", round(sargan_statistic, 3), "\n")
Sargan 統計量： 0.558
> cat("p 值：", round(p_value, 4), "\n")
p 值： 0.4549
>
> # 在 5% 顯著性水平下檢驗
> if (p_value < 0.05) {
+   cat("在 5% 顯著性水平下，p 值 < 0.05，拒絕工具變量外生的假設，RANK 和 POS 可能無效。\n")
+ } else {
+   cat("在 5% 顯著性水平下，p 值 >= 0.05，無法拒絕工具變量外生的假設，RANK 和 POS 可能有效。\n")
+ }
在 5% 顯著性水平下，p 值 >= 0.05，無法拒絕工具變量外生的假設，RANK 和 POS 可能有效。
```

10.24 Consider the data file *mroz* on working wives. Use the 428 observations on married women who participate in the labor force. In this exercise, we examine the effectiveness of alternative standard errors for the IV estimator. Estimate the model in Example 10.5 using IV/2SLS using both *MOTHEREDUC* and *FATHEREDUC* as IV. These will serve as our baseline results.

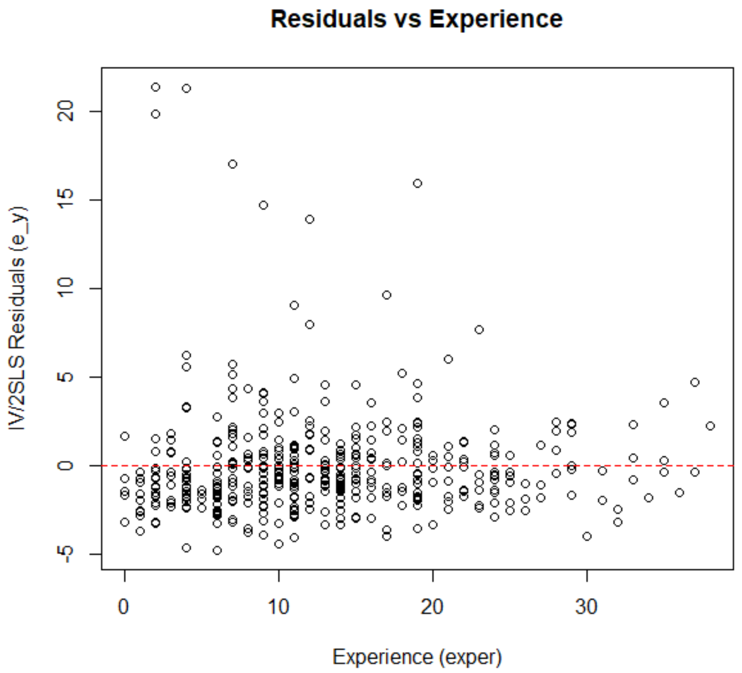
a. Calculate the IV/2SLS residuals, $\hat{e}_{IV}$. Plot them versus *EXPER*. Do the residuals exhibit a pattern consistent with homoskedasticity?

b. Regress $\hat{e}_{IV}^2$ against a constant and *EXPER*. Apply the NR^2 test from Chapter 8 to test for the presence of heteroskedasticity.

c. Obtain the IV/2SLS estimates with the software option for Heteroskedasticity Robust Standard Errors. Are the robust standard errors larger or smaller than those for the baseline model? Compute the 95% interval estimate for the coefficient of *EDUC* using the robust standard error.

d. Obtain the IV/2SLS estimates with the software option for Bootstrap standard errors, using $B = 200$ bootstrap replications. Are the bootstrap standard errors larger or smaller than those for the baseline model? How do they compare to the heteroskedasticity robust standard errors in (c)? Compute the 95% interval estimate for the coefficient of *EDUC* using the bootstrap standard error.

a.



b.

```
> cat("IV/2SLS 默認標準誤（假設同方差性）：\n")
IV/2SLS 默認標準誤（假設同方差性）：
> print(round(default_se, 4))
(Intercept)        educ       exper    exper_sq
     1.8622      0.1462      0.0625      0.0019
>
> # 計算穩健標準誤（HC1 類型）
> robust_se <- sqrt(diag(vcovHC(iv_model, type = "HC1")))
> cat("\nIV/2SLS 穩健標準誤（HC1）：\n")

IV/2SLS 穩健標準誤（HC1）：
> print(round(robust_se, 4))
(Intercept)        educ       exper    exper_sq
     2.1543      0.1618      0.0717      0.0019
```

標準誤比較（默認 vs 穩健）：

```
> comparison <- data.frame(Default_SE = defa
> print(round(comparison, 4))
            Default_SE Robust_SE Difference
(Intercept)     1.8622    2.1543    -0.2920
educ            0.1462    0.1618    -0.0156
exper           0.0625    0.0717    -0.0092
exper_sq        0.0019    0.0019     0.0000
```

判斷：

```
> if (any(abs(comparison$Difference) / default_se > 0.1)) {  # 如果差異超過 10%
+   cat("穩健標準誤與默認標準誤差異較大，IV 估計的標準誤與同方差性假設不一致，存在異方差性。\n")
+ } else {
+   cat("穩健標準誤與默認標準誤相近，IV 估計的標準誤與同方差性假設一致。\n")
+ }
穩健標準誤與默認標準誤差異較大，IV 估計的標準誤與同方差性假設不一致，存在異方差性。
```

c.

```
> print(se_comparison)
            Estimate Baseline_SE Robust_SE Increased_SE
(Intercept) -0.41651     1.86224   2.15428          Yes
educ         0.32824     0.14624   0.16180          Yes
exper        0.04617     0.06249   0.07173          Yes
exper_sq    -0.00069     0.00187   0.00190          Yes
> cat("Conclusion:\nRobust SEs are larger, indicating heteroskedasticit
y.\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: [0.0111, 0.6454]
```