**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?

```
Percentage of mothers with some college education:  10.09 %
Percentage of fathers with some college education:  10.76 %
```

**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*?

| | educ | MOTHERCOLL | FATHERCOLL |
|---|---|---|---|
| educ | 1.0000000 | 0.3370171 | 0.3193212 |
| MOTHERCOLL | 0.3370171 | 1.0000000 | 0.3674532 |
| FATHERCOLL | 0.3193212 | 0.3674532 | 1.0000000 |

MOTHERCOLL and FATHERCOLL show moderate positive correlations with educ, suggesting they are relevant instruments. Their binary form may also reduce bias from unobserved factors, making them preferable to motheredu and fatheredu.

**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 $\beta_j$ is the $j$th 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?

```
Call:
lm(formula = I(msft - riskfree) ~ I(mkt - riskfree), 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
I(mkt - riskfree) 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
```

$(r_{msft} - r_f) = 0.00325 + 1.20184 \cdot (r_m - r_f)$

The estimated beta for Microsoft is 1.20, indicating that it is riskier than the market portfolio during this period. Since the beta is significantly greater than 1, Microsoft stock shows higher sensitivity to market movements. ✱

**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?

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

The instrument RANK is highly significant in the first-stage regression, with an $R^2$ of 0.91 and an F-statistic of 1858. Therefore, RANK can be regarded as a strong instrument. ✱

**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?

```
Call:
lm(formula = I(msft - riskfree) ~ I(mkt - riskfree) + resid_v,
    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
I(mkt - riskfree) 1.278318   0.126749  10.085   <2e-16 ***
resid_v          -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
```

In the augmented CAPM model, the residual $\hat{v}$ is not significant at the 1% level. Therefore, we fail to reject the null hypothesis of exogeneity and 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?

```
Call:
ivreg(formula = I(msft - riskfree) ~ I(mkt - riskfree) | 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
I(mkt - riskfree) 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
```

Using RANK as an instrument, the IV estimate of beta is 1.278, which is slightly higher than the OLS estimate of 1.202. Both estimates are highly significant, and the similarity suggests that the market return is likely exogenous, which aligns with the result from (c). *

**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?

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

In the first-stage regression using both RANK and POS as instruments, the joint F-statistic is 951.3 and the $R^2$ is 0.915. Both instruments are statistically significant, and the high F-statistic strongly suggests that we have adequately strong IVs. *

**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?

```
Call:
lm(formula = I(msft - riskfree) ~ I(mkt - riskfree) + resid_v_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
I(mkt - riskfree) 1.283118   0.126344  10.156   <2e-16 ***
resid_v_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
```

In the augmented model, the coefficient on the residual $\hat{v}_e$ is not significant at the 1% level. Therefore, we fail to reject the null hypothesis of exogeneity and conclude that the market return is exogenous at the 1% level. ✱

**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?

```
Call:
ivreg(formula = I(msft - riskfree) ~ I(mkt - riskfree) | 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
I(mkt - riskfree) 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
```

Using RANK and POS as instruments, the IV estimate of beta is 1.283, which is slightly higher than the OLS estimate of 1.202. Both estimates are highly significant. This result is consistent with our expectations, especially considering the Hausmen test in (f), which supported the exogeneity of the market return. ✱

**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.

```
[1] 0.5584634
[1] 0.45488
```

⇒ The Sargan test yields a test statistic of 0.558 with a p-value of 0.455, which is above the 5% significant level. Therefore, we fail to reject the null hypothesis and conclude that the surplus instrument (PUS) is valid. The instruments appears to be exogenous. ※