The cost of attending your college has once again gone up. Although you have been told that education is investment in human capital, which carries a return of roughly 10% a year, you (and your parents) are not pleased. One of the administrators at your university/college does not make the situation better by telling you that you pay more because the reputation of your institution is better than that of others. To investigate this hypothesis, you collect data randomly for 100 national universities and liberal arts colleges from the 2000–2001 U.S. News and World Report annual rankings. Next you perform the following regression

\[
\hat{\text{Cost}} = 7,311.17 + 3,985.20 \times \text{Reputation} - 0.20 \times \text{Size} \\
+ 8,406.79 \times D_{\text{priv}} - 416.38 \times D_{\text{libart}} - 2,376.51 \times D_{\text{religion}}
\]

\[
(2,058.63) \quad (664.58) \quad (0.13) \\
(2,154.85) \quad (1,121.92) \quad (1,007.86)
\]

\[R^2 = 0.72, \quad \text{SER} = 3,773.35\]

where \(\text{Cost}\) is Tuition, Fees, Room and Board in dollars, \(\text{Reputation}\) is the index used in U.S. News and World Report (based on a survey of university presidents and chief academic officers), which ranges from 1 ("marginal") to 5 ("distinguished"), \(\text{Size}\) is the number of undergraduate students, and \(D_{\text{priv}}, D_{\text{libart}},\) and \(D_{\text{religion}}\) are binary variables indicating whether the institution is private, a liberal arts college, and has a religious affiliation. The numbers in parentheses are heteroskedasticity-robust standard errors.

(a) Interpret the results and indicate whether or not the coefficients are significantly different from zero. Do the coefficients have the expected sign?

(b) What is the forecasted cost for a liberal arts college, which has no religious affiliation, a size of 1,500 students and a reputation level of 4.5? (All liberal arts colleges are private.)

(c) To save money, you are willing to switch from a private university to a public university, which has a ranking of 0.5 less and 10,000 more students. What is the effect on your cost? Is it substantial?

(d) What is the \(p\)-value for the null hypothesis that the coefficient on \(\text{Size}\) is equal to zero? Based on this, should you eliminate the variable from the regression? Why or why not?

(e) You want to test simultaneously the hypotheses that \(\beta_{\text{Size}} = 0\) and \(\beta_{D_{\text{libart}}} = 0\). Your regression package returns the \(F\)-statistic of 1.23. Can you reject the null hypothesis?

(f) Eliminating the \(\text{Size}\) and \(D_{\text{libart}}\) variables from your regression, the estimation regression becomes

\[
\hat{\text{Cost}} = 5,450.35 + 3,538.84 \times \text{Reputation} + 10,935.70 \times D_{\text{priv}} - 2,783.31 \times D_{\text{religion}};
\]

\[
(1,772.35) \quad (590.49) \quad (875.51) \quad (1,180.57)
\]
Why do you think that the effect of attending a private institution has increased now?

(g) You give a final attempt to bring the effect of Size back into the equation by forcing the assumption of homoskedasticity onto your estimation. The results are as follows:

\[
\hat{\text{Cost}} = 7,311.17 + 3,985.20 \times \text{Reputation} - 0.20 \times \text{Size} \\
\quad + 8,406.79 \times \text{Dpriv} - 416.38 \times \text{Dlibart} - 2,376.51 \times \text{Dreligion} \\
\quad (1,985.17) \quad (593.65) \quad (0.07) \\
\quad (1,423.59) \quad (1,096.49) \quad (989.23)
\]

\[
R^2 = 0.72, \quad SER = 3,682.02
\]

Calculate the \( t \)-statistic on the Size coefficient and perform the hypothesis test that its coefficient is zero. Is this test reliable? Explain.

(h) What can you say about causation in the above relationship? Is it possible that \( \text{Cost} \) affects \( \text{Reputation} \) rather than the other way around?

2) (Requires Appendix Material) Consider the following multiple regression model

\[
Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i
\]

You want to consider certain hypotheses involving more than one parameter, and you know that the regression error is homoskedastic. You decide to test the joint hypotheses using the rule-of-thumb \( F \)-statistics. For each of the cases below specify a restricted model and indicate how you would compute the \( F \)-statistic to test for the validity of the restrictions.

(a) \( \beta_1 = -\beta_2; \beta_3 = 0 \)

(b) \( \beta_1 + \beta_2 + \beta_3 = 1 \)

(c) \( \beta_1 = \frac{\beta_2}{\beta_3} \)

(d) \( \beta_1 = -\beta_2; \beta_3 = 0 \)
3) Write the following four restrictions in the form $R\beta = r$, where the hypotheses are to be tested simultaneously.

$$\beta_3 = 2\beta_5,$$
$$\beta_1 + \beta_2 = 1,$$
$$\beta_4 = 0,$$
$$\beta_2 = -\beta_6.$$

Can you write the following restriction $\beta_2 = -\frac{\beta_2}{\beta_1}$ in the same format? Why not?
1) (a) An increase in reputation by one category, increases the cost by roughly $3,985. The larger the size of the college/university, the lower the cost. An increase of 10,000 students results in a $2,000 lower cost. Private schools charge roughly $8,406 more than public schools. A school with a religious affiliation is approximately $2,376 cheaper, presumably due to subsidies, and a liberal arts college also charges roughly $416 less. There are no observations close to the origin, so there is no direct interpretation of the intercept. Other than perhaps the coefficient on liberal arts colleges, all coefficients have the expected sign, although that coefficient is not significantly different from zero. All other coefficients are statistically significant at conventional levels, with the exception of the size coefficient, which carries a $t$-statistic of 1.54, and hence is not statistically significant at the 5% level (using a one-sided alternative hypothesis).

(b) $32,935.

(c) Roughly $12,400. Since over the four years of education, this implies approximately $50,000, it is a substantial amount of money for the average household.

(d) Using a one-sided alternative hypothesis, the $p$-value is 6.2 percent. Variables should not be eliminated simply on grounds of a statistical test. The sign of the coefficient is as expected, and its magnitude makes it important. It is best to leave the variable in the regression and let the reader decide whether or not this is convincing evidence that the size of the university matters.

(e) The critical value for $F_{2, \infty}$ is 3.00 (5% level) and 4.61 (1% level). Hence you cannot reject the null hypothesis in this case.

(f) Private institutions are smaller, on average, and some of these are liberal arts colleges. Both of these variables had negative coefficients.

(g) Although the coefficient would be statistically significant in this case, the test is unreliable and should not be used for statistical inference. There is no theoretical suggestion here that the errors might be homoskedastic. Since the standard errors are quite different here, you should use the more reliable ones, i.e., the heteroskedasticity-robust.

(h) It is very possible that the university president and chief academic officer are influenced by the cost variable in answering the U.S. News and World Report survey. If this were the case, then the above equation suffers from simultaneous causality bias, a topic that will be covered in a later chapter. However, this poses a serious threat to the internal validity of the study.
2) (a) The restricted model is \( Y_i = \beta_0 + \beta_2 (X_2 - X_3) + u_i \) and the rule-of-thumb \( F \)-statistic would be

\[
F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})}{SSR_{\text{unrestricted}} / (n - 3 - 1) / 2}.
\]

(b) \( (Y_i - X_{3i}) = \beta_0 + \beta_1 (X_2 - X_3) + \beta_2 (X_2 - X_3) + u_i \) and the rule-of-thumb \( F \)-statistic would be

\[
F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})}{SSR_{\text{unrestricted}} / (n - 3 - 1) / 1}.
\]

(c) This is not a linear restriction. Hence you cannot use the \( F \)-test to test for its validity.

(d) \( Y_i = \beta_0 + (\beta_1 X_2 + X_3) + u_i \) and the rule-of-thumb \( F \)-statistic would be

\[
F = \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})}{SSR_{\text{unrestricted}} / (n - 3 - 1) / 2}
\]

\[
\begin{pmatrix}
0 & 0 & 0 & 1 & 0 & -2 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\beta_5 \\
\beta_6
\end{pmatrix} =
\begin{pmatrix}
0 \\
1 \\
0 \\
0
\end{pmatrix}
\]

3) The restriction \( \beta_3 = -\frac{\beta_2}{\beta_1} \) cannot be written in the same format because it is nonlinear.