uncorrelated. Thus, if you are interested in , it is best to leave out of the regression if it is correlated with X_1."

6.11 (Requires calculus) Consider the regression model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

for $i = 1, \ldots, n$. (Notice that there is no constant term in the regression.) Following analysis like that used in Appendix 4.2:

a. Specify the least squares function that is minimized by OLS.

b. Compute the partial derivatives of the objective function with respect to b_1 and b_2.

c. Suppose $\sum_{i=1}^{n} X_{1i} X_{2i} = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^{n} X_{1i} Y_i / \sum_{i=1}^{n} X_{1i}^2$.

d. Suppose $\sum_{i=1}^{n} X_{1i} X_{2i} \neq 0$. Derive an expression for $\hat{\beta}_1$ as a function of the data (Y_i, X_{1i}, X_{2i}), $i = 1, \ldots, n$.

e. Suppose that the model includes an intercept: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Show that the least squares estimators satisfy $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2$.

f. As in (e), suppose that the model contains an intercept. Also suppose that $\sum_{i=1}^{n} (X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2) = 0$. Show that $\hat{\beta}_1 = \sum_{i=1}^{n} (X_{1i} - \overline{X}_1)(Y_i - \overline{Y}) / \sum_{i=1}^{n} (X_{1i} - \overline{X}_1)^2$. How does this compare to the OLS estimator of β_1 from the regression that omits X_2?

Empirical Exercises

E6.1 Using the data set **TeachingRatings** described in Empirical Exercises 4.2, carry out the following exercises.

a. Run a regression of *Course_Eval* on *Beauty*. What is the estimated slope?

b. Run a regression of *Course_Eval* on *Beauty*, including some additional variables to control for the type of course and professor characteristics. In particular, include as additional regressors *Intro*, *OneCredit*, *Female*, *Minority*, and *NNEnglish*. What is the estimated effect of *Beauty* on *Course_Eval*? Does the regression in (a) suffer from important omitted variable bias?

c. Estimate the coefficient on *Beauty* for the multiple regression model in (b) using the three-step process in Appendix 6.3 (the Frisch–Waugh theorem). Verify that the three-step process yields the same estimated coefficient for *Beauty* as that obtained in (b).

d. Professor Smith is a black male with average beauty and is a native English speaker. He teaches a three-credit upper-division course. Predict Professor Smith's course evaluation.

E6.2 Using the data set **CollegeDistance** described in Empirical Exercise 4.3, carry out the following exercises.

a. Run a regression of years of completed education (*ED*) on distance to the nearest college (*Dist*). What is the estimated slope?

b. Run a regression of *ED* on *Dist*, but include some additional regressors to control for characteristics of the student, the student's family, and the local labor market. In particular, include as additional regressors *Bytest*, *Female*, *Black*, *Hispanic*, *Incomehi*, *Ownhome*, *DadColl*, *Cue80*, and *Stwmfg80*. What is the estimated effect of *Dist* on *ED*?

c. Is the estimated effect of *Dist* on *ED* in the regression in (b) substantively different from the regression in (a)? Based on this, does the regression in (a) seem to suffer from important omitted variable bias?

d. Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and $\overline{R}^2$. Why are the R^2 and $\overline{R}^2$ so similar in regression (b)?

e. The value of the coefficient on *DadColl* is positive. What does this coefficient measure?

f. Explain why *Cue80* and *Swmfg80* appear in the regression. Are the signs of their estimated coefficients (+ or −) what you would have believed? Interpret the magnitudes of these coefficients.

g. Bob is a black male. His high school was 20 miles from the nearest college. His base-year composite test score (*Bytest*) was 58. His family income in 1980 was \$26,000, and his family owned a home. His mother attended college, but his father did not. The unemployment rate in his county was 7.5%, and the state average manufacturing hourly wage was \$9.75. Predict Bob's years of completed schooling using the regression in (b).

h. Jim has the same characteristics as Bob except that his high school was 40 miles from the nearest college. Predict Jim's years of completed schooling using the regression in (b).

E6.3 Using the data set **Growth** described in Empirical Exercise 4.4, but excluding the data for Malta, carry out the following exercises.

a. Construct a table that shows the sample mean, standard deviation, and minimum and maximum values for the series *Growth*, *TradeShare*, *YearsSchool*, *Oil*, *Rev_Coups*, *Assassinations*, *RGDP60*. Include the appropriate units for all entries.

b. Run a regression of *Growth* on *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations* and *RGDP60*. What is the value of the coefficient on *Rev_Coups*? Interpret the value of this coefficient. Is it large or small in a real-world sense?

c. Use the regression to predict the average annual growth rate for a country that has average values for all regressors.

d. Repeat (c) but now assume that the country's value for *TradeShare* is one standard deviation above the mean.

e. Why is *Oil* omitted from the regression? What would happen if it were included?

APPENDIX

6.1 Derivation of Equation (6.1)

This appendix presents a derivation of the formula for omitted variable bias in Equation (6.1). Equation (4.30) in Appendix 4.3 states that

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{6.16}$$

Under the last two assumptions in Key Concept 4.3, $(1/n)\sum_{i=1}^{n}(X_i - \overline{X})^2 \xrightarrow{p} \sigma_X^2$ and $(1/n)\sum_{i=1}^{n}(X_i - \overline{X})u_i \xrightarrow{p} \text{cov}(u_i, X_i) = \rho_{Xu}\sigma_u\sigma_X$. Substitution of these limits into Equation (6.16) yields Equation (6.1).