and consistent? Explain. Do you think that $E(u_i|X_{1i}, X_{2i})$ depends on X_2? Will the OLS estimator of β_2 provide an unbiased and consistent estimate of the causal effect of transferring to a new school (that is, being a newly-enrolled student)? Explain.

Empirical Exercises

E7.1 Use the data set **CPS08** described in Empirical Exercise 4.1 to answer the following questions.

a. Run a regression of average hourly earnings (*AHE*) on age (*Age*). What is the estimated intercept? What is the estimated slope?

b. Run a regression of *AHE* on *Age*, gender (*Female*), and education (*Bachelor*). What is the estimated effect of *Age* on earnings? Construct a 95% confidence interval for the coefficient on *Age* in the regression.

c. Are the results from the regression in (b) substantively different from the results in (a) regarding the effects of *Age* and *AHE*? Does the regression in (a) seem to suffer from omitted variable bias?

d. Bob is a 26-year-old male worker with a high school diploma. Predict Bob's earnings using the estimated regression in (b). Alexis is a 30-year-old female worker with a college degree. Predict Alexis's earnings using the regression.

e. 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)?

f. Are gender and education determinants of earnings? Test the null hypothesis that *Female* can be deleted from the regression. Test the null hypothesis that *Bachelor* can be deleted from the regression. Test the null hypothesis that both *Female* and *Bachelor* can be deleted from the regression.

g. A regression will suffer from omitted variable bias when two conditions hold. What are these two conditions? Do these conditions seem to hold here?

E7.2 Using the data set **TeachingRatings** described in Empirical Exercise 4.2, carry out the following exercises.

a. Run a regression of *Course_Eval* on *Beauty*. Construct a 95% confidence interval for the effect of *Beauty* on *Course_Eval*.

b. Consider the various control variables in the data set. Which do you think should be included in the regression? Using a table like Table 7.1, examine the robustness of the confidence interval that you constructed in (a). What is a reasonable 95% confidence interval for the effect of *Beauty* on *Course_Eval*?

E7.3 Use the data set **CollegeDistance** described in Empirical Exercise 4.3 to answer the following questions.

a. An education advocacy group argues that, on average, a person's educational attainment would increase by approximately 0.15 year if distance to the nearest college is decreased by 20 miles. Run a regression of years of completed education (*ED*) on distance to the nearest college (*Dist*). Is the advocacy groups' claim consistent with the estimated regression? Explain.

b. Other factors also affect how much college a person completes. Does controlling for these other factors change the estimated effect of distance on college years completed? To answer this question, construct a table like Table 7.1. Include a simple specification [constructed in (a)], a base specification (that includes a set of important control variables), and several modifications of the base specification. Discuss how the estimated effect of *Dist* on *ED* changes across the specifications.

c. It has been argued that, controlling for other factors, blacks and Hispanics complete more college than whites. Is this result consistent with the regressions that you constructed in part (b)?

E7.4 Using the data set **Growth** described in Empirical Exercise 4.4, but excluding the data for Malta, carry out the following exercises.

a. Run a regression of *Growth* on *TradeShare*, *YearsSchool*, *Rev_Coups*, *Assassinations*, and *RGDP60*. Construct a 95% confidence interval for the coefficient on *TradeShare*. Is the coefficient statistically significant at the 5% level?

b. Test whether, taken as a group, *YearsSchool*, *Rev_Coups*, *Assassinations*, and *RGDP60* can be omitted from the regression. What is the p-value of the F-statistic?