10.2 The labor supply of married women has been a subject of a great deal of economic research. Consider the following supply equation specification

$$HOURS = \beta_1 + \beta_2 WAGE + \beta_3 EDUC + \beta_4 AGE + \beta_5 KIDSL6 + \beta_6 KIDS618 \\ + \beta_7 NWIFEINC + e$$

where *HOURS* is the supply of labor, *WAGE* is hourly wage, *EDUC* is years of education, *KIDSL6* is the number of children in the household who are less than six years old, *KIDS618* is the number between 6 and 18 years old, and *NWIFEINC* is household income from sources other than the wife's employment.

(a) Discuss the signs you expect for each of the coefficients.

(b) Explain why this supply equation cannot be consistently estimated by least squares regression.

(c) Suppose we consider the woman's labor market experience *EXPER* and its square, $EXPER^2$, to be instruments for *WAGE*. Explain how these variables satisfy the logic of instrumental variables.

(d) Is the supply equation identified? Explain.

(e) Describe the steps (not computer commands) you would take to obtain *2SLS* estimates.

10.6 The 500 values of x, y, z_1, and z_2 in *ivreg2.dat* were generated artificially. The variable $y = \beta_1 + \beta_2 x + e = 3 + 1 \times x + e$.

(a) The explanatory variable x follows a normal distribution with mean zero and variance $\sigma_x^2 = 2$. The random error e is normally distributed with mean zero and variance $\sigma_e^2 = 1$. The covariance between x and e is 0.9. Using the algebraic definition of correlation, determine the correlation between x and e.

(b) Given the values of y and x, and the values of $\beta_1 = 3$ and $\beta_2 = 1$, solve for the values of the random disturbances e. Find the sample correlation between x and e and compare it to your answer in (a).

(c) In the same graph, plot the value of y against x, and the regression function $E(y) = 3 + 1 \times x$. Note that the data do not fall randomly about the regression function.

(d) Estimate the regression model $y = \beta_1 + \beta_2 x + e$ by least squares using a sample consisting of the first $N = 10$ observations on y and x. Repeat using $N = 20$, $N = 100$, and $N = 500$. What do you observe about the least squares estimates? Are they getting closer to the true values as the sample size increases, or not? If not, why not?

(e) The variables z_1 and z_2 were constructed to have normal distributions with means zero and variances one, and to be correlated with x but uncorrelated with e. Using the full set of 500 observations, find the sample correlations between z_1, z_2, x, and e. Will z_1 and z_2 make good instrumental variables? Why? Is one better than the other? Why?

(f) Estimate the model $y = \beta_1 + \beta_2 x + e$ by instrumental variables using a sample consisting of the first $N = 10$ observations and the instrument z_1. Repeat using $N = 20$, $N = 100$, and $N = 500$. What do you observe about the *IV* estimates? Are they getting closer to the true values as the sample size increases, or not? If not, why not?

(g) Estimate the model $y = \beta_1 + \beta_2 x + e$ by instrumental variables using a sample consisting of the first $N = 10$ observations and the instrument z_2. Repeat using $N = 20$, $N = 100$, and $N = 500$. What do you observe about the *IV* estimates? Are they getting closer to the true values as the sample size increases, or not? If not, why not? Comparing the results using z_1 alone to those using z_2 alone, which instrument leads to more precise estimation? Why is this so?

(h) Estimate the model $y = \beta_1 + \beta_2 x + e$ by instrumental variables using a sample consisting of the first $N = 10$ observations and the instruments z_1 and z_2. Repeat using $N = 20$, $N = 100$, and $N = 500$. What do you observe about the *IV* estimates? Are they getting closer to the true values as the sample size increases, or not? If not, why not? Is estimation more precise using two instruments than one, as in parts (f) and (g)?