7.1 An economics department at a large state university keeps track of its majors' starting salaries. Does taking econometrics affect starting salary? Let SAL = salary in dollars, GPA = grade point average on a 4.0 scale, $METRICS = 1$ if student took econometrics, and $METRICS = 0$ otherwise. Using the data file *metrics.dat*, which contains information on 50 recent graduates, we obtain the estimated regression

$$\begin{array}{lcccc} \widehat{SAL} & = & 24200 + 1643GPA + 5033METRICS & & R^2 = 0.74 \\ (\text{se}) & & (1078) \quad (352) \quad (456) & & \end{array}$$

(a) Interpret the estimated equation.
(b) How would you modify the equation to see whether women had lower starting salaries than men? (Hint: Define an indicator variable $FEMALE = 1$, if female; zero otherwise.)
(c) How would you modify the equation to see if the value of econometrics was the same for men and women?

7.4 In the file *stockton.dat* we have data from January 1991 to December 1996 on house prices, square footage, and other characteristics of 4682 houses that were sold in Stockton, California. One of the key problems regarding housing prices in a region concerns construction of "house price indexes," as discussed in Section 7.2.4b. To illustrate, we estimate a regression model for house price, including as explanatory variables the size of the house ($SQFT$), the age of the house (AGE), and annual indicator variables, omitting the indicator variable for the year 1991.

$$\begin{aligned} PRICE = {} & \beta_1 + \beta_2 SQFT + \beta_3 AGE + \delta_1 D92 + \delta_2 D93 + \delta_3 D94 + \delta_4 D95 \\ & + \delta_5 D96 + e \end{aligned}$$

The results are as follows:

Stockton House Price Index Model

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	21456.2000	1839.0400	11.6671	0.0000
SQFT	72.7878	1.0001	72.7773	0.0000
AGE	−179.4623	17.0112	−10.5496	0.0000
D92	−4392.8460	1270.9300	−3.4564	0.0006
D93	−10435.4700	1231.8000	−8.4717	0.0000
D94	−13173.5100	1211.4770	−10.8739	0.0000
D95	−19040.8300	1232.8080	−15.4451	0.0000
D96	−23663.5100	1194.9280	−19.8033	0.0000

7.15 The data file *br2.dat* contains data on 1080 house sales in Baton Rouge, Louisiana, during July and August 2005. The variables are *PRICE* ($), *SQFT* (total square feet), *BEDROOMS* (number), *BATHS* (number), *AGE* (years), *OWNER* (=1 if occupied by owner; zero if vacant or rented), *POOL* (=1 if present), *TRADITIONAL* (=1 if traditional style; 0 if other style), *FIREPLACE* (=1 if present), and *WATERFRONT* (=1 if on waterfront).

(a) Compute the data summary statistics and comment. In particular, construct a histogram of *PRICE*. What do you observe?

(b) Estimate a regression model explaining $\ln(PRICE/1000)$ as a function of the remaining variables. Divide the variable *SQFT* by 100 prior to estimation. Comment on how well the model fits the data. Discuss the signs and statistical significance of the estimated coefficients. Are the signs what you expect? Give an exact interpretation of the coefficient of *WATERFRONT*.

(c) Create a variable that is the product of *WATERFRONT* and *TRADITIONAL*. Add this variable to the model and reestimate. What is the effect of adding this variable? Interpret the coefficient of this interaction variable, and discuss its sign and statistical significance.

(d) It is arguable that the traditional-style homes may have a different regression function from the diverse set of nontraditional styles. Carry out a Chow test of the equivalence of the regression models for traditional versus nontraditional styles. What do you conclude?

(e) Using the equation estimated in part (d), predict the value of a traditional style house with 2500 square feet of area, that is 20 years old, that is owner-occupied

at the time of sale, that has a fireplace, 3 bedrooms, and 2 baths, but no pool, and that is not on the waterfront.

7.16* Data on 1500 house sales from Stockton, California, are contained in the data file *stockton4.dat*. [Note: *stockton3.dat* is a larger version of the same data set, containing 2610 observations.] The houses are detached single-family homes that were listed for sale between October 1, 1996, and November 30, 1998. The variables are *PRICE* ($), *LIVAREA* (hundreds of square feet), *BEDS* (number of bedrooms), *BATHS* (number of bathrooms), *LGELOT* (= 1 if lot size is greater than 0.5 acres, zero otherwise), *AGE* (years), and *POOL* (= 1 if home has pool, zero otherwise).

(a) Examine the histogram of *PRICE*. What do you observe? Create the variable ln(*PRICE*) and examine its histogram. Comment on the difference.

(b) Estimate a regression of ln(*PRICE*/1000) on the remaining variables. Discuss the estimation results. Comment on the signs and significance of the variables *LIVAREA*, *BEDS*, *BATHS*, *AGE*, and *POOL*.

(c) Discuss the effect of large lot size on the selling price of a house.

(d) Introduce to the model an interaction variable *LGELOT***LIVAREA*. Estimate this model and discuss the interpretation, sign, and significance of the coefficient of the interaction variable.

(e) Carry out a Chow test of the equivalence of models for houses that are on large lots and houses that are not.