Assignment # 4

Econometrics 616

1. **4.4.** The model

$$Y = \alpha_1 + \gamma_2 E_2 + \gamma_3 E_3 + u$$

is estimated by OLS, where E_2 and E_3 are dummy variables indicating membership of the second and third educational classes. Show that the OLS estimates are

$$\begin{bmatrix} a_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 - \bar{Y}_1 \\ \bar{Y}_3 - \bar{Y}_1 \end{bmatrix}$$

where $\bar{Y}_i$ denotes the mean value of Y in the ith educational class.

See Johnston & DiNardo Page 137

You Can use Computer!

2. **4.5.** Using the data in Table 4.4 estimate the specification

$$Y = \alpha_1 E_1 + \alpha_2 E_2 + \alpha_3 E_3 + \beta_2 S_2 + u$$

and estimate the resultant mean values for Table 4.5. Compare your results with the values given in the text and comment.

Repeat the exercise for the specification

$$Y = \mu + \alpha_1 E_1 + \alpha_2 E_2 + \beta_2 S_2 + u$$

3. **4.6.** Using the data of Table 4.4 estimate a specification of your own choosing without a constant term but with appropriate dummy variables to allow for interaction effects. Calculate the resultant version of Table 4.6 and compare with the results in the text.

4. The usual two-variable linear model is postulated, and a sample of 20 observations is drawn from an urban area and another sample of 10 observations from a rural area. The sample information in raw form is summarized as follows:

Urban

$$X'X = \begin{bmatrix} 20 & 20 \\ 20 & 25 \end{bmatrix} \qquad X'y = \begin{bmatrix} 10 \\ 20 \end{bmatrix} \qquad y'y = 30$$

Rural

$$X'X = \begin{bmatrix} 10 & 10 \\ 10 & 20 \end{bmatrix} \qquad X'y = \begin{bmatrix} 8 \\ 20 \end{bmatrix} \qquad y'y = 24$$

Test the hypothesis that the same relationship holds in both urban and rural areas.

5. Suppose $S = \alpha + \beta Ed + \phi IQ + \eta Ex + \lambda Sex + \delta DF + \theta DE + \varepsilon$ where S is salary, Ed is years of education, IQ is IQ level, Ex is years of on-the-job experience, Sex is one for males and zero for females, DF is one for French-only speakers and zero otherwise, DE is one for English-only speakers and zero otherwise. Given a sample of N individuals who speak only French, only English or are bilingual:

(a) Explain how you would test for discrimination against females (in the sense that *ceteris paribus* females earn less than males).

(b) Explain how you would measure the payoff to someone of becoming bilingual given that his or her mother tongue is (i) French, (ii) English.

(c) Explain how you would test the hypothesis that the two payoffs of the preceding question are equal.

(d) Explain how you would test the hypothesis that a French-only male earns as much as an English-only female.

(e) Explain how you would test if the influence of on-the-job experience is greater for males than for females.

6. Assume that the true regression model is of the form

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + u_i \qquad (1)$$

If the regression

$$Y_i = \beta_1^* + \beta_2^* X_{2i} + v_i \qquad (2)$$

is run, what can you say about the direction of the bias of the slope coefficient (b_2^*).

7. Consider two alternative regressions

1. $y = \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \epsilon.$
2. $y = \alpha_1 \quad + \alpha_2 D_2 + \alpha_3 D_3 + \alpha_4 D_4 + \epsilon.$

The variables are quarterly dummy variables [see (8-1)]. There are equal numbers of observations in each quarter. By computing $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ for the two models, obtain the precise formulas for the least squares coefficients in the two cases. Prove that if there were a term βx in the two equations, the least squares estimates of β would be identical.

<Skip this>

Greene P. 347-8

8. In a classic study of the American automobile market, Griliches (1961b) reports the following regression results:

$$\begin{aligned}\ln \text{Price} = {} & 6.4 + 0.056H + 0.249W + 0.023L + 0.010V + 0.023T \\ & + 0.090A + 0.088P + 0.109B + 0.157C - 0.044D_1 - 0.015D_2 \\ & + 0.019D_3 + 0.044D_4 + 0.044D_5 + 0.023D_6,\end{aligned}$$

where 1954 is the base year and

H = advertised horsepower, in hundreds;
W = shipping weight, in thousands of pounds;
L = overall length, in tens of inches;
V = 1 if the car has a V8 engine and 0 otherwise;
T = 1 if the car is a hard top, 0 if a convertible:
A = 1 if it has an automatic transmission, 0 if not;
P = 1 if power steering is standard, 0 if not;
B = 1 if power brakes are standard, 0 if not;
C = 1 if the car is a "compact," 0 if not;
D_1 = 1 for 1955 model car, 0 otherwise;
D_2 = 1 for 1956 model car, 0 otherwise;
D_3 = 1 for 1957 model car, 0 otherwise;
D_4 = 1 for 1958 model car, 0 otherwise;
D_5 = 1 for 1959 model car, 0 otherwise;
D_6 = 1 for 1960 model car, 0 otherwise.

(The value of 6.4 for the constant term is assumed here. The author did not report the constant terms in his results.)

The regression is estimated using a large sample of prices of cars for the years 1954 to 1960. Since the dependent variable is in logarithms and the coefficients are close to zero, they may be treated as percentage changes. For example, the model predicts that if everything else were held constant, an increase of 100 horsepower increased the cost of a car by 5.6 percent, whereas from 1954 to 1956, the price of a particular car fell by 1.5 percent.

(a) As a benchmark, suppose that an average car in 1954 had 141 horsepower, weighed 3452 pounds, and was 205 inches long (from Griliches, Table 1). Suppose, as well, that in 1954, 25 percent of the cars had V8s, 75 percent were hard tops, 20 percent had automatic transmissions, 10 percent had power steering, 5 percent had power brakes, and 10 percent of the cars sold called be called "compact." (All these are hypothetical; Griliches gives no figures for these.) What does the model predict for the median price of a car? (The prediction would not be the mean. Griliches reports a *geometric* mean of all prices for 1954 of $2360.)

(b) Everything else held constant, how much more would it cost to buy a car with a V8 and power steering in 1955 than without these two features in 1954?

(c) Holding all other features constant, how much did the price of a car rise from 1956 to 1957. From 1956 to 1959?

(d) Suppose that putting a hard top on a car instead of a convertible always added exactly 100 pounds to the weight of the car. How would this affect the interpretation of the coefficients on W and T in the regression?

(e) The R^2 in the regression above was 0.922, using 570 observations. If the R^2 in the regression without $D_1, \ldots$ were 0.919, would the results support the hypothesis that variation in the mix of options and other features was entirely responsible for the variation in car prices in the 7-year period?

(f) Suppose that the preceding data had also been used to compute seven separate regressions, one for each year, with the following results:

	1954	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*	*All*
Observations	65	55	87	95	103	87	78	570
e′e	104	88	206	144	199	308	211	1425

Test the hypothesis that the same coefficients apply in each year. [Note: The model with the pooled data allows a separate constant for each year. Note as well that you may need an approximate value for the F distribution with large degrees of freedom.]