

EE 325 STATA SESSION PRACTICE II

Multiple regression, Dummy variables, and Heteroscedasticity

1. Table 8.8 The Cobb-Douglas Production Function for the Mexican Economy, 1955 – 1974

$$\ln GDP_t = \beta_1 + \beta_2 \ln Labor_t + \beta_3 \ln Capital_t \quad (1.1)$$

$$\ln \left(\frac{GDP_t}{Labor_t} \right) = \gamma_1 + \gamma_2 \ln \left(\frac{Capital_t}{Labor_t} \right) \quad (1.2)$$

Is the Mexican economy characterized by constant returns to scale over the sample period?

2. Table 9.1 Public school teachers' salaries (Dollars) by geographical region. Consider and estimate the following model with the dummy variable:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

Where Y = (average) salary of public school teacher in state i

$D_{2i} = 1$ if the state is in the Northeast or North Central
= 0 otherwise

$D_{3i} = 1$ if the state is in the South
= 0 otherwise

t = time

$D = 1$ for observations in 1970-1981
= 0 otherwise

Find the mean salary of public school teacher in the South and the mean salary of public school teacher in the Northeast or North Central.

3. Table 8.9 Savings and Personal Disposable income (billions of dollars), United States, 1970-1995.

- a. Given the data in the table, estimate the following linear savings function using personal disposable income

$$\text{Time period 1970-1981: } Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$$

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}$ $n_2 = 14$

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_{3t}$ $n_1 + n_2 = 26$

- b. On the basis of the Chow test that there was a difference in the regression of savings on income between the two periods. Consider and estimate the following model with the dummy variable:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

Where Y = Savings

X = Personal disposable income

t = time

D = 1 for observations in 1982-1995

= 0 otherwise

- (i) Test the coefficients individually statistically significant at the 5 percent level? From this test, how would you describe the difference in the two regressions (coincident regression, parallel regression, concurrent regression, dissimilar regression)?
- (ii) Write down the mean personal savings function for 1970-1981 and the mean personal savings function for 1982-2005.
4. Table 9.7 provides data on a sample of 114 workers in an industrial town in southern India in 1990. The variables are defined as follows:
- WI = weekly wage income in rupees
- Age = age in years
- D_{sex} = 1 for male workers and 0 for female workers
- DE_2 = a dummy variable taking a value of 1 for workers with an education level up to primary
- DE_3 = a dummy variable taking a value of 1 for workers up to a secondary level of education
- DE_4 = a dummy variable taking a value of 1 for workers with higher than secondary education

DPT = a dummy variable taking a value of 1 for workers with permanent jobs and a value of 0 for temporary workers

We estimate the following regression model:

(1)

$$\ln WI_i = \beta_1 + \beta_2 AGE_i + \beta_3 D_{SEX} + \beta_4 DE_2 + \beta_5 DE_3 + \beta_6 DE_4 + \beta_7 DPT + u_i$$

(2)

$$\ln WI_i = \lambda_1 + \lambda_2 AGE_i + \lambda_3 D_{SEX} + \lambda_4 DE_2 + \lambda_5 DE_3 + \lambda_6 DE_4 + \lambda_7 D_{SEX} DE_2 + \lambda_8 D_{SEX} DE_3 + \lambda_9 D_{SEX} DE_4 + \lambda_{10} DPT + u_i$$

(3)

$$\ln WI_i = \gamma_1 + \gamma_2 AGE_i + \gamma_3 D_{SEX} + \gamma_4 D_{SEX} DE_2 + \gamma_5 D_{SEX} DE_3 + \gamma_6 D_{SEX} DE_4 + \gamma_7 DPT + u_i$$

5. Table 11.1 The relationship between compensation and employment size

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Is the Homogenous assumption valid? If the assumption on Homogenous does not hold, perform GLS estimation. Assume population variance is known.

Methods of Weighted Least Squares

$$\frac{Y_i}{\sigma_i} = \beta_1 * \left(\frac{1}{\sigma_i}\right) + \beta_2 * \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

6. Table 11.5 R&D Expenditure, Sales, and Profits in 14 Industry Groupings in the United States, 2005 Since the cross-sectional data presented in this table are quite heterogenous, in a regression of R&D on sales, heteroscedasticity is likely. The regression results are as follows:

$$R\&D_i = \beta_1 + \beta_2 Sales_i + u_i$$