Econometrics Lab 4

Linear Regression with Qualitative Information

1. Testing Cobb-Douglas Production Function We use the datasets dairy.dat from Lab 3. In the data, the first 9 observations are United States (1954), Canada (1954), New Zealand (1955/56), Australia (1955/1956), Denmark (1954), Norway (1954), United Kingdom (1951), Ireland (1953), and Puerto Rico (1952). These are developed countries. The last 7 countries are Colombia (1953), Mexico (1951), Argentina (1950), El Salvador (1951), Southern Rhodesia (1952), Ceylon (1952), and Japan (1953). These are developing countries at that time.

An objection of the conclusion in Lab 3 (2) is that the developed and the developing countries should not have the same production technology. To account for this, we can use a dummy variable called *developed*, $developed = 1$ if the country belongs to the first group and $developed = 0$ if it belongs to the second group.

(1) Construct the new variable *developed* by hand in the data file.

(2) Estimate the following regression and discuss your results (goodness of fit, significance of estimates, economic meanings, etc.).

$$\log(V/L) = \beta_0 + \beta_1 \log w + \beta_2 developed + u. \tag{1}$$

(3) Test the following hypothesis and discuss your results,

$$\mathrm{H}_0: \ \beta_2 = 0 \quad \mathrm{H}_1: \ \beta_2 > 0.$$

2. Income Determination. In this exercise we use cgss05.csv, the dataset we use in Lab 2 on the question of income determination. We use the following variables.

$income$	yearly income
edu	years of schooling
$expr$	number of years after graduation
$female$	dummy for females
$rural$	dummy for rural residence status (Hukou)
$prv01 \sim prv28$	dummy for provinces

In this exercise we consider the problem of wage discrimination in China based on gender and residence and the problem of income inequality among provinces.

Before we begin, note that the variable *income* include bequest, interest and dividend payment, etc., as well as wage income. Due to the unavailability of data, we assume that only an insignificant portion of income is derived from other than wages. Based on anecdotal observations, this assumption should be a reasonable one.

(1) Estimate the following model. Read significance tests from the results.

$$\log(income) = \beta_0 + \beta_1 edu + \beta_2 expr + \beta_3 female + \beta_4 rural + u. \tag{2}$$

(2) Test whether there exists significant discrimination against females, controlling for education, experience, and Hukou.

(3) Test whether there exists significant discrimination against workers with rural Hukou, controlling for education, experience, and gender.

(4) Test whether the same increase in education brings more income to a female than to a male. You need to write a new model and do tests on that.

(5) Test whether the female is even more disadvantaged if she holds the rural Hukou. You need to write a different model on which the test can be conducted.

(6) It is well known that wages are widely different across different provinces in China. Write a model to consider that difference. Beware of the problem of multicollinearity.

(7) Test whether there exist significant income differentials across different provinces, controlling for education, experience, gender, and Hukou.

3. Labor Participation for Women in the US In this exercise we use the dataset mroz87.txt, which is from Mroz (1987) study of labor supply behavior of married women. The definitions of variables are inside the data file. We use the following variables,

lfp	A dummy variable = 1 if woman worked in 1975, else 0
wa	Wife's age
we	Wife's educational attainment, in years
$faminc$	Family income, in 1975 dollars
$kl6$	Number of children less than 6 years old in household
$k618$	Number of children between ages 6 and 18 in household
mtr	The marginal tax rate facing the wife

(1) Define a binary variable called "kids", which takes 1 if there are children under 18 in the family and 0 otherwise. And estimate a linear probability model,

$$lfp = \beta_0 + \beta_1 wa + \beta_2 wa^2 + \beta_3 we + \beta_4 faminc + \beta_5 kids + u. \tag{3}$$

Remember to choose White procedure to obtain heteroscedasticity-robust standard errors.

(2) Discuss the marginal effects of age, education, family income, and having kids.

(3) For a woman at age 60 with 20 years' of education, if she has no kids and her family income is 100000, what is the probability for her to participate the labor force? Does the number make sense?

(4) Estimate a probit model with the same variables as in Equation (3). Calculate the probability of being in labor force for the woman in (3). What if she has a kid under 18?

(5) For a woman with average age, education, family income, and with no kids, calculate the marginal effect of education on her probability of labor force participation. What if she has kids?

(6) For the "average" woman in (5), calculate the marginal effect of age on her probability of labor force participation.

(7) By now you may find that family income does not have a significant influence on women's labor participation decision. From now on we do not include it in our model. Instead, you may conjecture that the marginal tax rate probably has some influence. Test this conjecture.

(8) Estimate the logit model with the same variables in (7). Calculate the marginal effect of mtr on the probability of being in labor force for a woman with average age, education, mtr, and with no kids. What if she has kids?