

Applied Econometrics, Heterogeneity in Treatment Effect

Before you begin the exercises:

- Download the data set “racewage_educ.dta” from the course website.
- Note: the outcome measure of interest is **re78**, real earnings in 1978, after the program is over. The variable with the randomly assigned treatment status is **treated**.

Exercise A: Allowing effect of training program to differ by two race categories

A1-What is the average effect of treatment to the NSW training program on earnings in 1978 for men?

A2-What is the average effect of treatment to the NSW training program on earnings in 1978 for black men? Hint: you can estimate same equation as in A1 but restricting sample to blacks (e.g., `reg y x if black==1`).

A3-What is the average effect of treatment to the NSW training program on earnings in 1978 for non-black men? Hint: you can estimate same equation as in A1 but restricting sample to blacks (e.g., `reg y x if black==0`).

A4-Can you reject the null hypothesis that the program treatment effect is the same for blacks and whites at the 5% significance level? Hint: See Lecture 6 Slide 3 Part A for the single equation you could estimate by OLS with the whole sample to assess this. You will need to form a new variable which is the interaction between treated and black.

Exercise B: Allowing effect of training program to differ by two income categories

B1-Form a variable indicating if an individual has below-average earnings in the baseline (use **re75** as the baseline income). Do this by summarizing **re75**, seeing the mean for that variable, and then forming the desired variable (e.g., after typing “`sum re75`” type “`gen lowearn=re75<r(mean)`”). You can verify that this did what you wished by typing “`tabulate re75 lowearn`” or “`sort lowearn`” then “`by lowearn: sum re75`”.

B2-What is the average effect of treatment to the NSW training program on earnings in 1978 for men with low baseline earnings? Hint: you can estimate same equation as in A1 but restricting sample to `lowearn==1` individuals.

B3- What is the average effect of treatment to the NSW training program on earnings in 1978 for men with low baseline earnings? Hint: you can estimate same equation as in A1 but restricting sample to `lowearn==0` individuals.

B4- Can you reject the null hypothesis that the program treatment effect is the same for blacks and whites at the 5% significance level? Hint: See what you did in A4.

Exercise C: Allowing effect of training program to differ by continuous income measure

Just as in Exercise B, we are interested in exploring treatment effect heterogeneity by income. However, instead of measuring income with a binary variable indicating above or below mean, we can measure it other ways. Here, we use the income variable as is instead of breaking it up into categories and using dummies for each category.

C1-Form a new variable that is the interaction between treated and re75, call this interact1. Now type: “reg re78 treated interact1 re75”.

- a. The coefficient for interact1 gives you the extra treatment effect for each additional dollar of baseline income.
- b. The coefficient for treated gives you the effect of treatment for someone with zero re75. (Sometimes nobody has zero value for the continuous variable (e.g., in Lecture 6 bank account example, nobody was age 0) and so this treatment effect is really not interesting and you wish to calculate the treatment effect at other values of the continuous variable such as the mean. But, actually in this example, this treatment effect for individuals with no earnings in 1975 may be of interest, you can see by doing “tab re75” that quite a lot of these individuals have 0 earnings in 1975 and this would be the effect of treatment for them. But for the sake of practicing demeaning a continuous baseline variable and obtaining estimates where the coefficient for the treatment dummy gives the effect for someone with the average value for that baseline variable, we will go ahead with C2.) If you wanted to use the coefficients you obtained here to get the effect of treatment for someone with the mean earnings in 1975, then type: “display f + g*h” where f=OLS coefficient for treated, g=OLS coefficient for interact1 and h is the mean of re75 (or you can type this into your calculator to perform the calculation).

C2-Form a new variable that is the re75 variable with the mean removed (type “sum re75” then type “gen dmre75=re75-r(mean)” and you can verify that dmre75 has mean zero, so when you are calculating the effect when this demeaned variable equals zero you are really calculating the effect when the original variable is at the mean). Now form the interaction between treated and dmre75, call this interact2. Now type: “reg re78 treated interact2 re75”.

- a. The coefficient for interact2 gives you the extra treatment effect for each additional dollar of baseline income. Verify it is the same as what you got in part C1a above.
- b. The coefficient for treated gives you the effect of treatment for someone with zero re75dm, i.e., for someone with mean re75. Verify that this is the same as what you calculated in part C1b above for the effect on individuals with average re75.

This ends part 1 of Lab 4 on using the regression framework to allow for (and test for) heterogeneity in treatment effects by some baseline characteristic. At 11am, Professor Chin will begin part 2 on power tests.