Introduction to Econometrics
Summer 2019 (11 Week): Hunter College
Problem Set 4: Due July 17, 2019
R. Tasnim

Answer the following questions using Stata. You have to submit the log file with your commands, and output.

1. Use the dataset Wage2.dta to answer the following questions.

   (a) Estimate the following model:

   $$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + u$$

   (b) Holding other factors fixed, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant?

   (c) Add the variables $exper^2$ and $tenure^2$ to the equation and test the joint hypothesis: $H_0 : \beta_{exper^2} = 0$ and $\beta_{tenure^2} = 0$.

   (d) Start with the original model, but now allow wages to differ across four groups of people: married and black, married and nonblack, single and black, and single and nonblack. Compare the wages of these four groups of people using both:

      i. an interaction term
      ii. creating dummy variables to represent the groups

2. Use the dataset GPA2.dta to answer the following questions.

   (a) Consider the following regression:

   $$ColGpa = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 Athelete + u$$

   where $ColGpa$ is cumulative college GPA, $hsize$ is size of high school graduating class (in hundreds), $hsperc$ is academic percentile in graduating class, sat is combined SAT score, $female$ is a binary variable for gender (= 1 if female and = 0 for male), and $athlete$ is a binary variable (= 1 if athlete and = 0 if not an athlete). Before running the regression, what are your expectations of the coefficients for sat and $athlete$ in this regression. Why? (Hint: keep in mind that we have quite a few controlling variables).

   (b) Estimate the regression in (a). Interpret the coefficient for $athlete$. Is this coefficient statistically significant?

   (c) Drop sat from the regression. Interpret the coefficient for $athlete$. Discuss why the coefficient for $athlete$ here is different than the one found in (b).

   (d) In the model in (a), allow the effect of being an athlete on $ColGpa$ differ by gender. Test the null hypothesis that holding everything else constant, there is no difference between female athletes and female non-athletes.

   (e) In the model in (a), allow for the effect of sat on $ColGpa$ to differ by gender. Test the null hypothesis that the effect of sat differs by gender.

3. Use the dataset loanapp.dta to answer the following questions.

   (a) Consider the following regression:

   $$approve = \beta_0 + \beta_1 white + u$$

   where *approve* is a binary variable (= 1 if mortgage loan was approved and = 0 if the loan was denied), *white* is a binary variable (= 1 if the applicant is white and = 0 if the applicant is nonwhite (hispanic or black)). Controlling for an individual's characteristics, if there is racial discrimination, what do you expect the sign of β_1 to be?

   (b) Estimate the regression in (a) using a linear probability model. Interpret the coefficient for *white*. Is it statistically significant? Is it large or small?

   (c) As controls, add the variables *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat*1, *mortlat*2, and *vr*. What happens to the coefficient for *white*. Is there still evidence of discrimination against nonwhites?

   (d) Reestimate the regression in (c) but with only the variables which are statistically significant at 10% significant level. Does your estimate for *white* change substantially? What does this say about omitted variable bias?

   (e) Allow the effect of race to interact with the variable measuring other obligations as a percentage of income (*obrat*). Is the interaction term significant?

   (f) In the regression in (c), what is the difference in predicted probability of *approve* between white and nonwhite applicants when they are both male ($male = 1$)?

   (g) Estimate the regression in (c) using a logit model. What is the difference in predicted probability of *approve* between white and nonwhite applicants when they are both male ($male = 1$) and married ($married = 1$)?

   (h) Estimate the regression in (c) using a probit model. What is the difference in predicted probability of *approve* between white and nonwhite applicants when they are both male ($male = 1$) and married ($married = 1$)?

   (i) Compare the results for the last three questions.