ECOM20001: Econometrics 1

Tutorial 11: Logarithmic Regression and Interactions

A. Getting Started

Please create a Tutorial11 folder on your computer, and then go to the LMS site for ECOM 20001 and download the following files into the Tutorial11 folder:

- tute11.R
- tute11_cps.csv

The first file is the R code for tutorial 11, the second file is the .csv file that contains the dataset for the tutorial.[1] The dataset has the following 5 variables:

- year: year individual was randomly surveyed; either 1992 or 2012
- ahe: individual's average hourly earnings (in real terms, 2012=100)
- bachelor: equals 1 if individual has a bachelor degree, 0 otherwise
- female: equals 1 if individual is female, 0 otherwise
- age: age of the individual at time of survey

In total, the dataset contains this information for 15,052 individuals in the U.S.

B. Go to the Code

With the R file downloaded into your Tutorial11 folder, you are ready to proceed with the tutorial. Please go to the tute11.R file to continue with the tutorial.

Note: This tutorial is slightly longer than previous tutorials and includes up to three separate presentations. It is expected that the tutors will work through the first two presentations in the tutorial, and that the third presentation material will be for students to individually work through with the tutorial code and solutions.

---

[1] The reference for these data is the Current Population Survey (CPS) which is collected by the U.S. Department of Labor Statistics and provides individual-level data on the population, employment, and earnings. It is constructed from randomly sampling the U.S. population. For details, see https://www.census.gov/programs-surveys/cps.html

C. Questions

Having worked through the tute11.R code and graphs, please answer the following:

Logarithmic Regressions

1. Construct three new logarithmic variables using the dataset:
   - log(ahe) = logarithm of ahe
   - log(age) = logarithm of age
   - d1992 = dummy variable equals 1 if the year is 1992, and is 0 otherwise
2. Run the following 4 regressions and interpret the sign, magnitude and statistical significance of any independent variables that involve age:
   - Regression 1: Linear
     - Dependent variable: ahe
     - Independent variables: age, bachelor, female, d1992
   - Regression 2: Linear-Log
     - Dependent variable: ahe
     - Independent variables: log_age, bachelor, female, d1992
   - Regression 3: Log-Linear
     - Dependent variable: log_ahe
     - Independent variables: age, bachelor, female, d1992
   - Regression 4: Log-Log
     - Dependent variable: log_ahe
     - Independent variables: log_age, bachelor, female, d1992

Interactions

3. Construct two new interactive variables using the dataset:
    - female_age = female x age
    - female_bachelor = female x bachelor
4. Run the following 2 regressions, comment on the statistical significance of the coefficients involving age, female, or bachelor, and compute the following partial effects involving age, female, or bachelor:
    - Regression 5:
        - Dependent variable: ahe
        - Independent variables: age, bachelor, female, female_age, d1992
        - Compute the following partial effects:
            - Partial effect of being one year older if male
            - Partial effect of being one year older if female
            - Partial effect of being female if 25 years old
            - Partial effect of being female if 30 years old
            - Partial effect of having a bachelor's degree
    - Regression 6:
        - Dependent variable: ahe
        - Independent variables: age, bachelor, female, female_bachelor, d1992
        - Compute the following partial effects:
            - Partial effect of being one year older
            - Partial effect of having a bachelor's degree if female
            - Partial effect of having a bachelor's degree if male
            - Partial effect of being female if you have a bachelor's degree
            - Partial effect of being female if you not have a bachelor's degree
5. Using Regression 5 in question 4 compute the partial effect of being female on ahe at age=28 and also report the 95% confidence interval (CI) for the partial effect.

Combining Logarithmic Regression and Interactions

6. Construct a new logarithmic variable that involves the logarithm of age interacted with the female dummy variable:
   - female_log_age = female x log(age)
7. Run the following 2 regressions, comment on the statistical significance of the coefficients involving age or female, and compute the following partial effects involving age or female:
   - Regression 7:
     - Dependent variable: log_ahe
     - Independent variables: age, bachelor, female, female_age, d1992
     - Compute the following partial effects:
       - Partial effect of being one year older if you are male
       - Partial effect of being one year older if you are female
       - Partial effect of being female if you are 25 years old
       - Partial effect of being female if you are 30 years old
   - Regression 8:
     - Dependent variable: log_ahe
     - Independent variables: log_age, bachelor, female, female_log_age, d1992
     - Compute the following partial effects (elasticities because of log-log form):
       - Partial effect / elasticity of being one year older if you are male
       - Partial effect / elasticity of being one year older if you are female
8. Using Regression 8 in question 7, compute the elasticity of ahe with respect to age for females. Also report the 95% CI for this elasticity.