ECOM20001: Econometrics 1

Tutorial 11: Combining 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. This code builds directly upon tute10.R from the previous week in combining logarithmic and interactive models. We also consider another example of a difference-in-differences model.

---

[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:

Combining Logarithmic Regression and Interactions

1. Construct a new logarithmic variable that involves the logarithm of age interacted with the female dummy variable:
   - female_log_age = female x log(age)
2. 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 1:
     - 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 2:
     - 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
3. Using Regression 2, compute the elasticity of ahe with respect to age for females. Also report the 95% CI for this elasticity.

Combining Interactions: Difference-in-Differences

Recall that in Tutorial 10 we saw a difference-in-differences (DiD) model with Regression 6 of that tutorial which quantified the differential impact of a Bachelor's degree on earnings between females and males. Here, we consider another DiD model, this time looking at whether the impact of a Bachelor's degree on earnings has had a differential impact over time.

4. First, construct the following interactive variable (where recall is created previously in the tute11.R code d2012 = 1- d1992)
    - bachelor_d2012 = bachelor x d2012
5. Now run the following regression and interpret the magnitude and statistical significance of the coefficient estimate on bachelor_d2012:[2]
    - Regression 3:
        - Dependent variable: log_ahe
        - Independent variables: bachelor, female, age, d2012, bachelor_d2012

[2] We could have alternatively developed this analysis identically using the d1992 variable instead. With the specification we work with, the difference-in-differences coefficient on bachelor_d2012 has a direct interpretation of the incremental value of a bachelor's degree on earnings in 2012 relative to its value in 1992. But we could have obtained an identical analysis working with d1992 and analogously-defined bachelor_d2012 instead.