Problem Set #4

ECON 833, Prof. Jason DeBacker
Due Tuesday, October 5, 10:05 a.m.

This problem requires you to define a statistical objective function in Python and then use SciPy to find the parameter values that minimizer that function.

The Stata data file `PS4_data.dta` provides you with the PSID used by Heathcoate, Perri, and Violante (*Review of Economic Dynamics*, 2010). The raw data are available here. I've modified this only to deflate the labor incomes of heads and spouses (`hlabinc` and `wlabinc`) to 2005$.

You need to do the following:

1. Select only male heads of household who are between 25 and 60 years of age and earn wages > \$7/hr.

2. Create indicator and continuous variables as necessary (see model below).

3. Estimate the following model via a Maximum Likelihood Estimator separately for $t = 1971, 1980, 1990, 2000$:

$$ln(w_{i,t}) = \alpha + \beta_1 Educ_{i,t} + \beta_2 Age_{i,t} + \beta_3 Age_{i,t}^2 + \beta_4 Black_{i,t} + \beta_5 Hispanic_{i,t} + \beta_6 OtherRace_{i,t} + \varepsilon_{i,t},$$

where:

- $w_{i,t}$ = wage of individual i in survey year t
- $Educ_{i,t}$ = education in years
- $Age_{i,t}$ = age in years
- $Black_{i,t}, Hispanic_{i,t}, OtherRace_{i,t}$ = dummy variables for race = Black, Hispanic, Not $\in$ {White, Black, Hispanic}.

4. Interpret the coefficient β_1. How do the returns to education change over time in these data?

You may submit your answers as a Jupyter Notebook, where you show your code and provide any text necessary to document your code and answer (4). Alternatively, you might have a set of `.py` files with your code and a pdf (compiled in TeX) in which you put your answers. Regardless of the method you choose to use, please organize the code and your documentation clearly. And please write out the equation for the likelihood function that you estimate. You will submit your problem set by pushing the files to your GitHub repository that you created from forking the repository for this class. You will put this and all other problem sets in the path `/CompEcon_Fall2021/ProblemSets/ProblemSet4`.

The following variable definitions are provided for your benefit (the PSID provides further documentation):

- `hlabinc` = annual labor income of the head

- `hannhrs` = annual hours of the head

- `hsex` = gender of the head (1=Male, 2=Female)

- `hrace` = race of the head (1=White, 2=Black, 3=Native American, 4=Asian/Pacific Islander, 5=Hispanic, 6,7=Other)

- `age` = age of the head

- `hyrsed` = years of education of the head

HINT: This model is linear. So you can check your MLE against an OLS estimator to confirm your results. See this QuantEcon notebook for a short tutorial on linear regressions in Python. Also, you may want a bounded optimizer e.g., try the "L-BFGS-B" and "SLSQP" methods in `scipy.optimize.minimize()` .