Due: May. 1

Econ 5253 - Spring 2025 **Problem Set 12** **beginning of class**

This problem set will give you practice with implementing some of the structural modeling techniques we discussed in class. A portion of the problem set will be closely related to Problem Set 7, which introduced methods for imputing missing data. As with the previous problem sets, you will submit this problem set by pushing the document to *your* (private) fork of the class repository. You will put this and all other problem sets in the path `/DScourseS25/ProblemSets/PS12/` and name the file `PS12_LastName.*`. Your OSCER home directory and GitHub repository should be perfectly in sync, such that I should be able to find these materials by looking in either place. Your directory should contain at least three files:

- `PS12_LastName.R` (you can also do this in Python or Julia if you prefer)
- `PS12_LastName.tex`
- `PS12_LastName.pdf`

1. Type `git pull origin master` from your OSCER `DScourseS25` folder to make sure your OSCER folder is synchronized with your GitHub repository.

2. Synchronize your fork with the class repository by doing a `git fetch upstream` and then merging the resulting branch.

3. Install the following R packages (if you don't already have them installed):
   - `sampleSelection`
   - `tidyverse`
   - `modelsummary`

4. Using R or Python, load the file `wages12.csv` (located in the current directory) in as a data frame. This data set contains information on $\approx 2,250$ women who were working in the US in 1988. The variables should be self-explanatory. `exper` refers to how long (in years) each woman has worked at an employer, and `hgc`, which refers to how many years of schooling each woman has completed. `union` is a binary variable indicating whether or not the woman is currently holding a union job, and `kids` is an indicator for wither the woman has at least one children living at home.

5. Format the `college`, `married`, and `union` variables as factors.

6. Use `modelsummary` to produce a summary table of this data frame. Include it in your LaTeXwriteup and discuss whether the results make sense. At what rate are log wages missing? Do you think the `logwage` variable is most likely to be MCAR, MAR, or MNAR?

7. As in PS7, perform the following imputation methods for missing `logwage` observations. **For each imputation method, estimate the following linear regression model:**

   $$logwage_i = \beta_0 + \beta_1 hgc_i + \beta_2 union_i + \beta_3 college_i + \beta_4 exper_i + \beta_5 exper_i^2 + \eta_i$$

   Our coefficient of interest is β_1 which can be interpreted as the returns to schooling (where schooling is thought of as an "investment" in "human captial"—we are curious what the Return On Investment is).

   - estimate the regression using only complete cases (i.e. do listwise deletion on the log wage variable ... this assumes log wages are Missing Completely At Random)
   - perform mean imputation to fill in missing log wages
   - use the `sampleSelection` package to correct for possible non-random missingness in the wages. To do this, follow these steps:
     - create a new variable called `valid` which flags log wage observations that are not missing
     - recode the log wage variable so that *invalid* observations are now equal to 0
     - estimate the Heckman selection (or "Heckit") model by issuing the following code:

```
selection(selection = valid ~ hgc + union + college + exper + married + kids,
          outcome = logwage ~ hgc + union + college + exper + I(exper^2),
          data = wagedata, method = "2step")
```

   Once you have finished all of this, use `modelsummary` to create one regression table which has the estimates of the first three regression models. Include this table in your .tex writeup. The true value of $\hat{\beta}_1 = 0.091$. Comment on the differences of $\hat{\beta}_1$ across the models. What patterns do you see? What can you conclude about the veracity of the various imputation methods?

8. Using the same data, estimate a probit model of preferences for working in a union job. To do this, use the `glm` function. The utility model has the following form:

   $$u_i = \alpha_0 + \alpha_1 hgc_i + \alpha_2 college_i + \alpha_3 exper_i + \alpha_4 married_i + \alpha_5 kids_i + \varepsilon_i$$

   where u_i is the utility of working in a union job (relative to working in a non-union job).

9. Assess the impact of a counterfactual policy in which wives and mothers are no longer allowed to work in union jobs. To do so, follow these steps: (if you don't remember how to do this, there is code in the GitHub repository in the file `example.R` in the `Structural` folder)

   - compute predicted probabilities of the model
   - change the coefficients on `married` and `kids` to equal zero
   - compute predicted probabilities associated with the new parameter values
   - compare the average of each set of predicted probabilities

   Do you think that the model you estimated above is realistic? Why or why not?

10. Compile your .tex file, download the PDF and .tex file, and transfer it to your cloned repository on OSCER. There are many ways to do this; you may ask an AI chatbot or simply drag-and-drop using VS Code. Do **not** put these files in your fork on your personal laptop; otherwise git will detect a merge conflict and that will be a painful process to resolve.

11. You should turn in the following files: .tex, .pdf, and any additional scripts (e.g. .R, .py, or .jl) required to reproduce your work. Make sure that these files each have the correct naming convention (see top of this problem set for directions) and are located in the correct directory (i.e. `~/DScourseS25/ProblemSets/PS12`).

12. Synchronize your local git repository (in your OSCER home directory) with your GitHub fork by using the commands in Problem Set 2 (i.e. `git add`, `git commit -m "message"`, and `git push origin master`). More simply, you may also just go to your fork on GitHub and click the button that says "Fetch upstream." Then make sure to pull any changes to your local copy of the fork. Once you have done this, issue a `git pull` from the location of your other local git repository (e.g. on your personal computer). Verify that the PS12 files appear in the appropriate place in your other local repository.