Advanced Econometrics: Homework 3

Instructions:

- The contact person for all agenda regarding this homework is Mathieu Petit, e-mail: mathieu.petit@fsv.cuni.cz
- Please form groups of three students. If you have trouble finding colleagues, write me an e-mail, and I will match you with others having the same problem.
- **Submit one file for the whole homework. Submit by e-mail to mathieu.petit@fsv.cuni.cz with the following subject:**
  "AdvEcon HW3 2023: Group surname1, surname2, surname3" .
- Deadline for submissions is **Friday, December 22, 2023, 23:59.** Any late submissions will be awarded zero points.
- Your solution should have a form of Jupyter Notebook with R source-code, other forms of submission are penalised by half the points. Code should be properly commented, interpretations of results as well as theoretical derivations should be written in markdown cells. This is the only file you need to send. If you prefer not to write formulas in LaTeX, you can send PDF with your derivations and interpretations in additional file and R code in Jupyter Notebook.
- Please, be concise, but remember to include and explain all important steps.
- If you have any questions concerning the homework, do contact me by mail and we can set up a consultation. Do it rather sooner than later, I won't give any consultation concerning the homework after December 15.

Problem 1:
(3 points)

Simulate $n = 500$ data points from the linear model

$$y_i = \alpha + \beta x_i + \epsilon_i$$

,
where $x_1, ..., x_n$ are observations from a random sample $X_1, ..., X_n \sim \mathcal{N}^n(14, 25)$ and $\epsilon|x \sim \mathcal{N}^n\left(0, x^{2\gamma}\right)$. Let's assume that the model parameters are $\alpha = 4$, $\beta = 1$, and $\gamma = 0.5$. The goal in this problem is to correctly estimate those parameters.

Remember to use **set.seed()** to make your results replicable. Recall that the parameters of the normal distribution are given in the form $\mathcal{N}(\mu, \sigma^2)$, not $\mathcal{N}(\mu, \sigma)$.

a) Estimate the parameters of the linear model $y = \alpha + \beta x + \epsilon$ on the simulated data using the OLS estimator. Interpret the results. Do you expect any of the OLS assumptions to be violated? If so, make the corresponding tests, and interpret the results.

b) Reestimate the model using GLS using your knowledge about the Ω matrix. State which form of the variance-covariance matrix Ω. Also, please state the form of the weighting matrix Ω^{-1}. Comment on the results from GLS regression.

c) Estimate FGLS model for heteroscedasticity of the form $\sigma_i^2 = h(x_i)$ (recall the food expenditure example from seminar 7).

d) In the OLS model, estimate standard errors using White heteroscedasticity consistent estimator. Compare White's standard errors to those from OLS, GLS, and FGLS.

Problem 2:
(3 points)

Please use the data in the **plm** R package with the R commands

```
library(plm)
data("Wages", package = "plm")
Wag <- pdata.frame(Wages, index=595)
```

to answer following questions. Estimate your models based on specification

$$lwage_{it} = \beta_0 + \beta_1 exp_{it} + \beta_2 exp_{it}^2 + \beta_3 sex_{it} + \beta_4 ed_{it} + \beta_5 south_{it} + \beta_6 smsa_{it} + \beta_7 black_{it} + u_{it},$$

where i is indicated by *id* variable, t is indicated by *time* variable. The dependent variable is natural logarithm of wage, *exp* indicates working experience, *ed* indicates years of education, *smsa* is a dummy variable for individuals living in urban areas, *sex* is a dummy variable indicating female workers.
Your task is to find out, whether education is a significant determinant of wages. Please use the standard panel estimation methods (Pooled OLS, Fixed Effects, and Random effects models), and perform all the necessary tests. For each estimator, state the conditions under which it is valid. What are your conclusions? What additional estimation would you suggest?

Problem 3:
(2 points)

Please use the data in the **plm** R package with the R commands

```
library(plm)
data("Wages", package = "plm")
Wag <- pdata.frame(Wages, index=595)
```

, and create a subsample containing all individuals for the second year of observation (i.e., *time*=2). Hence, you should estimate a model of following specification on a cross-sectional subsample

$$ln(wage_i) = \beta_0 + \beta_1 ed_i + \beta_2 smsa_i + \beta_3 black_i + u_i.$$

Test for the group specific heteroskedasticity based on the outcome of BLACK variable. Construct an FGLS estimator efficient in presence of such relationship. Remember Seminar 7.