Econometrics of Panel Data
Homework #3
Due to 31st January, 2021, 8.00 P.M.

General information: homework should be submitted electronically (via email using the SGH email address) and contain two files: pdf with solution and do-file (or other code, e.g. R script) that allows to replicate results. Please title your mail *[SGH] Econometrics of Panel Data. Homework 3* .

Exercise 1 (Probability of unemployment). Use the Stata built-in dataset `nlswork` and create the following indicator variable:

$$\mathcal{U}_{it} = \begin{cases} 0 & \text{if} \quad wks\_ue = 0 \\ 1 & \text{if} \quad wks\_ue > 0 \end{cases} \tag{1}$$

(a) Discuss difference in economic interpretation between $\mathcal{U}_{it}$ and wks_ue.

(b) Estimate the parameters of the pooled logit model in which $\mathcal{U}_{it}$ is the dependent variable. As explanatory variables use age, squared age, race indicators, dummy variables on occupations, industries, region, place of residence and collage graduate. Discuss which factors affect the probability of unemployment.

(c) Estimate the parameters of the RE logit model and compare estimates to the previous point. Discuss differences and similarities. Is there any evidence in favor of presence of unit-specific effects?

(d) Use now the FE logit regression. Compare results with previous points. Try to explain differences.

(e) Calculate and interpret odds ratio for the race indicator variables for all estimates.

(f) Calculate and interpret marginal effects of age for all estimates.

(g) Add time effects and compare the results to the previous point. Is there any trend effect? How the estimated time effects can be interpreted economically.

Exercise 2 (The COVID-19 pandemic and the effectiveness of lockdown). Consider the following model:

$$covid\_cases_{it} = \beta_{0,i} + \beta_{1,i}stringency\_index_{it} + \varepsilon_{it} \tag{2}$$

where the variable $covid\_cases_{it}$ measures pandemic situation in the i-th country, t is the time period, $stringency\_index_{it}$ is the so-called stringency index and it measures policy responses that governments have taken to react the COVID-19 pandemic (higher value is related to a larger number of policies or stricter polices) and ε_{it} is the error term.

(a) In the previous homework it has been assumed that all parameters are identical across countries. Discuss why this assumption seems to be unrealistic.

(b) Download once the latest data related to the COVID-19 pandemic and the stringency index. You can use the Our World in Data dataset ([link to the csv. file], [description of variables]). Choose new cases per million as proxy variable for $covid\_cases_{it}$ and estimate (2) with standard panel data estimators, i.e., the FE estimation. In addition, consider the mean group (MG) estimator. Compare results. Discuss differences between estimates and key assumptions employed in both methods.

(c) Use now pooled mean group (PMG) estimates . Compare results with MG estimates using the Hausman test. Is there any evidence in favor of homogeneity of parameters?

(d) Test whether there is cross-sectional dependence of residuals in both the MG and the PMG estimates. What are the consequences of the cross-sectional depended residuals?

(e) Add cross-sectional averages of outcome and explanatory variable. Interpret Hausman test comparing MG and PMG. Compare estimates with previous point and test cross-sectional estimates.

(f) Replicate previous point but use additionally several lags of cross-sectional averages.

(g) Based on the previous point select mean group estimates and perform analysis of cross-sectional variation in $\beta_{1,i}$. Use time invariant variables in regression analysis. Could we explain systematic differences in $\beta_{1,i}$?