Econometrics of Panel Data
Homework #1
Due to 27th November, 2020, 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 1* .

Exercise 1 (The relationship between crime and punishment)**.** Cornwell and Trumbul (1994) study the relationship between crime and punishment. Their dataset contains cross sectional data on 90 countries from 1981–1987. The subset of their data are in the file `crime_subset.dta`. Consider the following model:

$$\begin{aligned} LCRMRTE_{it} &= \alpha + \beta_1 LPRBARR_{it} + \beta_2 LPRBCONV_{it} + \beta_3 LPRBPRIS_{it} + \\ &= +\beta_4 LAVGSEN_{it} + \beta_5 LWMFG_{it} + u_{it} \end{aligned}$$

where (all variables are in logs):

$LCRMRTE_{it}$	– the crime rate
$LPRBARR_{it}$	– the probability of arrest,
$LPRBCONV_{it}$	– the probability of conviction,
$LPRBPRIS_{it}$	– the probability of a prison sentence,
$LAVGSEN_{it}$	– the average prison sentence,
$LWMFG_{it}$	– the average weekly wage in the manufacturing sector.

(a) What signs would you expect on the various coefficients? Why?

(b) Estimate the model using a pooled OLS. Discuss the results. Are they as you expected?

(c) Estimate the model using a fixed effects estimator. Interpret all coefficients. Are these results consistent with your expectations?

(d) Compare the estimates from (c) to the pooled regression. Explain why these estimates differ.

(e) Test the hypothesis that there are no cross-country differences.

(f) Discuss the significance of the estimated coefficients from the fixed effects estimators. Estimate once again the fixed effect estimator using robust standard errors. Compare the standard errors with those you obtained in (c). Explain the difference.

(g) Estimate the model using a random effect estimator. Interpret the estimated σ_u^2 and ρ.

(h) Compare the results from (g) with those you obtained in (c) and (b). Explain these differences/similarities.

(i) Perform Hausman's test for the estimates obtained in(c) and (g). Discuss the results.

(j) Add to the baseline specification dummy variables for the years 1982–1987 (denoted as *D*82-*D*87). Re-estimate the model using fixed and random effects estimators. Compare the estimated coefficients between the fixed and random effect model.

(k) Can you identify any trend effect? Test the joint significance of the period dummy variables in each model from (j).

(l) Based on the estimates obtained in (j), perform Hausman's test. Discuss the results. Which assumptions could be not satisfied? Should we rely on these results?

(m) Test the presence of the cross-country heterogeneity using the Lagrange multiplier statistic.

(n) In all above models, you've used an assumption that the variance-covariance matrix of error term is spherical. Is this assumption satisfied? Based on the residuals from regressions discussed in (j) provide some illustrative evidence on potential serial correlation, heteroskedasticity and cross-sectional dependence of the error term.

(o) Based on your investigation in (n) apply appropriate GLS estimation. Compare estimates ant their significance with results obtained in (j).

Exercise 2. Consider the following three-way RE model:

$$y_{its} = \alpha + \beta_1 x_{1its} + \ldots + \beta_k x_{kits} + u_{its}, \quad i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}, s \in \{1, \ldots, S\}, \quad (1)$$

where the index s stands for sector and the error term is as follows:

$$u_{its} = \mu_i + \lambda_t + \varphi_s + \varepsilon_{its}, \quad (2)$$

where $\mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$, $\lambda_t \sim \mathcal{N}\left(0, \sigma_\lambda^2\right)$, $\varphi_s \sim \mathcal{N}\left(0, \sigma_\varphi^2\right)$ and $\varepsilon_{its} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

(a) Derive the variance-covariance matrix of the error term. Give an intuitive explanation for this result.

(b) Can you generalize your results from the point (a) to the four-way RE model?

Exercise 3 (The COVID-19 pandemic and the effectiveness of lockdown)**.** Consider the following model:

$$covid\_cases_{it} = \beta_0 + \beta_1 stringency\_index_{it} + \varepsilon_{it} \quad (3)$$

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) What signs would you expect on the coefficient β_1. Why?

(b) Download 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 proxy variable for $covid\_cases_{it}$ and discuss this choice.

(c) Estimate the parameter of (3) using pooled regression as well as the RE and FE estimators. Remember about clustered standard errors. Are the estimates of β_1 in line with your expectations? Test the heterogeneity in the RE and FE models. Compare RE and FE models with the Hausman test.

(d) Choose at least two control variables that potentially affect the outcome variable and extend (3) by the selected variables. What sign would you expect on the related coefficients? Estimate the extended regression with pooled regression as well as the RE and FE estimators. Are the estimates of β_1 in line with your expectations? Test the heterogeneity in the RE and FE models. Compare RE and FE models with the Hausman test.

(e) Add time effects to the previous regressions and re-estimate parameters. Is there any trend effect? If so, try to visualize this effect for all models. Once again, are the estimates of β_1 in line with your expectations? Test the heterogeneity in the RE and FE models. Compare RE and FE models with the Hausman test.

(f) Discuss limitations of the above analysis and explain what is their impact on your analysis. Is specification of the considered model correct? Propose potential extension of your analysis. Apply your proposition if it is possible.