Econometrics of Panel Data
The retake exam session
Due to 18th February, 2021, 8.00 P.M.

General information: solutions 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. The retake exam session* .

Exercise 1. Consider the following gravity model of trade:

$$\ln EX_{ijt} = \beta_0 + \beta_1 \ln GDP_{it} + \beta_2 \ln GDP_{jt} + \beta_3 \ln dist_{ij} + \varepsilon_{ijt} \tag{1}$$

where $\ln EX_{ijt}$ is the logged gross exports from country i to the country j while $dist_{ijt}$ is the distance between them, GDP denotes the Gross Domestic Product and ε_{ijt} is the error term.

(i) Using the dataset `Gravity.dta` and estimate the underlying parameters parameters. Interpret all estimates.

(ii) Use clustered standard errors. Cluster the variance by: (i) importers, (ii) exporters, (iii) pairs consisting of importer and exports. Compare results. Which standard errors seem to be the most reliable?

(iii) Assume that the error term consists of two components: idiosyncratic and unit-specific. Provide economic interpretation for the latter one.

(iv) Estimate the random effects (RE) model for (1). Compare results with estimates from previous points. Interpret ρ.

(v) Perform LM test for variance of unit-specific individuals effects and interpret result.

(vi) Estimate the random effects (RE) model for (1). Compare results with previous estimates. Explain why it is not possible to obtain the FE estimates of β_3.

(vii) Interpret the results of poolability test.

(viii) Perform the Hausman test. Interpret result and try to explain economically this result.

(ix) Include time effects in (1). Namely, generate dummy variables for each year and replicate points, i.e., pooled regression, FE, RE and corresponding tests.

(x) Can you identify any trend effect? Test the joint significance of the period dummy variables in each model from previous point

(xi) Based on the FE estimates with time effects discuss to what an extent the variation in individual effects can explained by the (logged) distance.

(xii) Create dummy variable indicating that both importer and exporter belong to the European Union. Estimate the RE mode for (1) extended by this dummy variable and time effects. Interpret the estimates of the dummy variable.

(xiii) Using the Hausman-Taylor estimator reestimate (1) extended by dummy variable for the EU countries as well as time effects. Discuss which time-invariant variable can be endogenous and why. Compare results with previous point.

Exercise 2. Consider the following production function at the firm levels:

$$\ln Y_{it} = \beta_0 + \beta_1 \ln L_{it} + \beta_2 \ln K_{it} + u_{it}, \tag{2}$$

where Y_{it} is the production, L_{it} is the labor input, K_{it} is the capital stock and u_{it} is the error term.

(i) Use the file `chilean.dta` and estimate pooled regression. Use clustered standard errors. Interpret estimates and test the constant return to scale hypothesis.

(ii) Interpret economically the error term. Discuss the structure of the error term. How do you interpret the idiosyncratic and firm-specific components?

(iii) Estimate the random and fixed effects model for (2). Compare estimates with the Hausman test and interpret the estimates. Interpret the results of the poolability test.

(iv) Consider now dynamic model:

$$\ln Y_{it} = \beta_0 + \rho \ln Y_{it-1} + \beta_1 \ln L_{it} + \beta_2 \ln K_{it} + u_{it}, \tag{3}$$

where ρ is the persistence parameter and ρ is less in modulus than one. Discuss why the dynamic relationship can be appealing in this case.

(v) Use the Anderson-Hsiao estimator for the model (3). Interpret the estimate using short- and long-run multipliers.

(vi) Use the Arellano-Bond estimator (two step estimator with robust standard errors) for the model (3) Interpret the estimate using short- and long-run multipliers. Test serial correlation of the error term and overidentifying restrictions and interpret obtained results. Test the hypothesis about constant returns to scale.

(vii) Use the system GMM (Bond-Bover) estimator (two step estimator with robust standard errors) for the model (3). Interpret the estimate using short- and long-run multipliers. Test serial correlation of the error term and overidentifying restrictions and interpret obtained results. Test the hypothesis about constant returns to scale.

(viii) Replicate all previous calculations but extend previous regressions by time effects. Are your previous conclusion still valid. Is there any convincing evidence in favor of constant returns to scale?

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

$$covid\_cases_{it} = \beta_0 + \beta_1 stringency\_index_{it} + \varepsilon_{it} \tag{4}$$

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.

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

(ii) 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.

(iii) Convert the dataset to the weekly frequency. Estimate the parameter of (4) 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.

(iv) Choose at least two control variables that potentially affect the outcome variable and extend (4) 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.

(v) 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.

(vi) Consider the following model:

$$covid\_cases_{it} = \beta_0 + \rho_1 covid\_cases_{it-1} + \beta_1 stringency\_index_{it} + \varepsilon_{it} \tag{5}$$

in the previous points it was assumed that $\rho_1 = 0$. Discuss why this assumption seems to be unrealistic.

(vii) Using weekly dataset estimate (5) with standard panel data estimators, i.e., (i) the Anderson-Hsiao estimator, (ii) The Arellano-Bond estimator, and (iii) so called system GMM estimator. Compare the estimates in terms of the short- and long-run multipliers. For the GMM estimates report and discuss the result of test on serial correlation as well as the Sargan test. In the GMM estimation use two-step estimators and robust standard errors.

(viii) Extend the baseline dynamic model by twice lagged dependent variable:

$$covid\_cases_{it} = \beta_0 + \rho_1 covid\_cases_{it-1} + \rho_2 covid\_cases_{it-2} + \beta_1 stringency\_index_{it} + \varepsilon_{it} \tag{6}$$

and discuss when it is useful to estimate the parameter ρ_2.

(ix) Add previously, estimate (6) with standard panel data estimators, i.e., (i) the Anderson-Hsiao estimator, (ii) The Arellano-Bond estimator, and (iii) so called system GMM estimator. Compare the estimates in terms of the short- and long-run multipliers. In the GMM use two-step estimators and robust standard errors.

(x) Based on previous point discuss whether there is some evidence in favour of ρ_2 and based on this discussion choose one of the considered specification.

(xi) Are there any reasons to claim that the $stringency\_index_{it}$ is endogenous variable.

(xii) Set $stringency\_index_{it}$ as endogenous variable and reestimate (4) or (6). Compare estimates with results obtained in previous points (remember about test on serial correlation of the error term as well as Sargan test).

(xiii) Add time effects to estimation and compare results to the previous point.