Econometrics of Panel Data
Homework #1
Due to 14th 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 2* .

Exercise 1 (Demand for cigarettes)**.** Baltagi and Li (1986) study a demand for cigarettes. Consider the subset of their dataset that contains data on 46 US states from 1963-1982. The following variables are available:

- $state_{it}$ – the identification number of state
- $year_{it}$ – the year,
- $price_{it}$ – price per pack of cigarettes (in nominal terms),
- cpi_{it} – consumer price index (1983=100),
- $income_{it}$ – per capita disposable income (in nominal terms),
- $sales_{it}$ – cigarette sales in packs per capita.

(a) At the first stage, consider the following econometric model describing demand for cigarettes:

$$\ln C_{it} = \beta_0 + \beta_1 \ln Y_{it} + \beta_2 \ln P_{it} + \varepsilon_{it} \tag{1}$$

where C_{it} is the consumption of cigarettes, Y_{it} is the real disposable income per capita and P_{it} is the real price per pack of cigarettes. Why the log transformation is useful here?

(b) Based on the available dataset, propose and discuss the construction of variables that are able to identify empirically C_{it}, Y_{it} and P_{it}.

(c) What signs would you expect on β_1 and β_2? Why?

(d) Estimate the model using a pooled OLS, a RE estimator and FE estimator. Report all estimates in one table, interpret (remember that the variables are in logs) and discuss your results. Are they as you expected? Explain why these estimates differ.

(e) Test the hypothesis that there are no cross-state differences and test the presence of the cross-state heterogeneity using the Lagrange multiplier statistic. Do these results lead to the same conclusions?

(f) Perform Hausman's test for the estimates obtained in (d) for the RE and FE model. Interpret this result.

(g) Generate dummy variables for years 1964-1982 and add them to the baseline specification of the model. Reestimate the FE and RE model considered in (d) and, once again, discuss your estimates. Do they differ from the results obtained in point (d)? Test also the hypotheses considered in point (e) and discuss whether your conclusions from point (e) are still valid.

(h) Perform Hausman's test for the estimates obtained in (g)? Discuss this result.

(i) 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 (g) provide some illustrative evidence on potential serial correlation or heteroskedasticity of the error term.

(j) Apply the GLS estimation to the models considered in the point (g) and compare these estimates to the results presented in (g). Try to explain the differences in the estimates.

(k) Modify your model by including the lagged dependent variable:

$$\ln C_{it} = \beta_0 + \gamma \ln C_{it-1} + \beta_1 \ln Y_{it} + \beta_2 \ln P_{it} + \varepsilon_{it} \tag{2}$$

Why the above (dynamic) specification seems to be more reliable? Is the above specification supported by your previous conclusions? Why?

(l) Estimate the model from (k) by using a FE estimator with time dummy variables. Discuss and interpret estimates of short- and long-run elasticities. Are they reliable? Why?

(m) Estimate the model from (k) by using an IV estimator proposed by Anderson and Hsiao (1981). What is wrong with these estimates?

(n) Estimate the model from (k) by using a GMM estimator proposed by Arellano and Bond (1991). Discuss and interpret your estimates of short- and long-run elasticities. Test serial correlation of the error term of order two. Discuss the results of the Sargan's test.

(o) Estimate the model from (k) by using a (system) GMM estimator proposed by Bond and Bover (1998). Do your estimates differ from the results from points (n) and (l)? Discuss the results of the test for autocorrelation of the error term as well as the Sargan's test.

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

$$covid\_cases_{it} = \beta_0 + \rho_1 covid\_cases_{it-1} + \beta_1 stringency\_index_{it} + \varepsilon_{it} \tag{3}$$

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) In the previous homework it was assumed that $\rho_1 = 0$. 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 proxy variable for $covid\_cases_{it}$ and estimate (3) 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.

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

$$covid\_cases_{it} = \beta_0 + \rho_1 covid\_cases_{it-1} + \rho_2 covid\_cases_{it-2} + \beta_1 stringency\_index_{it} + \varepsilon_{it} \tag{4}$$

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

(d) Add previously, estimate (4) 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.

(e) 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.

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

(g) Set $stringency\_index_{it}$ as endogenous variable and reestimate (3) or (4). Compare estimates with results obtained in previous points (remember about test on serial correlation of the error term as well as Sargan test).

(h) Consider now aggregation of the analyzed dataset to weekly frequency. Discuss advantages and disadvantages of such transformation.

(i) Aggregate data to weekly frequency. And repeat previous calculation (at least from point (g)). Compare results.

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