Advanced Applied Econometrics
Homework
Due to: 6th June, 2026, 20.00

General information: solution 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] Advanced Applied Econometrics. Homework 2.*

Exercise 1. Consider the following ADL model for inflation:

$$\pi_t = \beta_0 + \sum_{i=1}^{P} \rho_p \pi_{t-i} + \beta_1 oil_t + \beta_2 food_t + \beta_3 neer_t + \beta_4 \tilde{y}_t + \beta_5 GSCPI_t + \varepsilon_t, \tag{1}$$

where π_t is the inflation rate, oil_t is price dynamics of oil, $food_t$ is the global inflation of food, $neer_t$ capture fluctuations in the nominal effective exchange rate , $\tilde{y}_t$ denotes the proxy of output gap, $GSCPI_t$is the global supply chains pressure index and ε_t is the error term.

(i) What signs would you expect on the coefficients. Explain.

(ii) Dataset `GermanInflation.dta` contains the monthly observations for Germany. Using this dataset estimate underlying parameters. Use the first difference of the logged HICP as measure of inflation. Take also the first differences of the logged oil prices, food price index, nominal effective exchange rate. In addition, as the $\tilde{y}_t$ use the first differences of the logged industrial production. After all transformations test stationarity of the series of the interest.

(iii) Assuming that $P = 1$ estimate parameters and compare with your initial expectations.

(iv) Based on available statistical tests determine the order of autoregressive part P that provides credible estimates. Compare with previous results in terms of the short- and long-run multipliers.

(v) Move to the VAR model and discuss the ordering of variables. Based on serial correlation tests select the number of lags and check key properties of estimates (normality of the error term, stability of VAR model).

(vi) Illustrate the reaction of inflation to exogenous shocks. Are they in line with your expectations? Sketch the FEVD for inflation and interpret this figure. Which shocks are the most important.

(vii) Consider now another strategy of transforming data. Namely, use original GSCPI data and the logged (non-stationary) series for the rest of economic variables. Estimate the VAR model with exogenous linear and quadratic deterministic trend. Compare the IRFs and FEVD to the previous case and try to explain differences.

(viii) Based on the regression results from (iv) provide historical decomposition of inflation rate since the beginning of 2020, however, provide the calculations for the annual inflation rate. To what extent the inflation rate in Germany can be explained by external factors? Which factors are predominant?

(ix) Replicate the previous point but with the VAR estimates. Are main findings the same?

Exercise 2. Consider the following gravity model of trade:

$$\ln EX_{ijt} = \beta_0 + \beta_1 \ln GDP_{it} + \beta_2 \ln GDP_{jt} + \beta_3 \ln dist_{ij} + \varepsilon_{ijt} \tag{2}$$

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

(ii) Using the dataset `Gravity.dta` estimate the underlying parameters parameters using pooled estimator, random effects (RE) estimator and fixed effects (FE) estimators. Interpret results and discuss differences.

(iii) Interpret ρ for the RE model. Perform LM test for variance of unit-specific individuals effects and interpret result.

(iv) Interpret the results of poolability test for the FE model.

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

(vi) Include time effects in (2). Namely, generate dummy variables for each year and replicate points, i.e., pooled regression, FE, RE and corresponding tests.

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

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

(ix) Create dummy variable indicating that both importer and exporter belong to the European Union. Estimate the RE model for (2) extended by this dummy variable and time effects. Interpret the estimates of the dummy variable.

(x) Using the Hausman-Taylor estimator reestimate (2) extended by dummy variable for the EU countries as well as time effects. Discuss which variable could be endogenous and why. Compare results with previous point.

Exercise 3. The birth weight is very important issue in health economics. In particular, a low birth weight of infants is important since its general effects are ranging from mortality to cognitive abilities in adulthood.

(i) Using the dataset `birthweight.dta` construct the following indicator variable:

$$lbw_i = \begin{cases} 1 & \text{if} \quad bwght_i < 2500 \\ 0 & \text{if} \quad bwght_i \geq 2500 \end{cases} \tag{3}$$

where $bwght_i$ is the birth weight of infant in grams. Discuss why above classification is better than the standard regression approach when the observed variance of the birth weight is explained.

(ii) Estimate the parameters of the logit regression in which lbw_i is the dependent variable. Use the remaining variable in the dataset as explanatory variables. Discuss which factors affect the probability of low weight of infants.

(iii) Calculate and interpret the odds ratio for the `smoke` and `omaps` variables.

(iv) Calculate and interpret the marginal effects of the `smoke` and `omaps` variables. In addition, show how the marginal effect of smoking is varying over the `omaps`.

(v) Based on the estimates provide in-sample classification of cases. How many cases are correctly classified? Compare this number with a fraction of infants with low birth weight. In addition, calculate how many cases with low birth weight are incorrectly classified.

(vi) However, the effect of low birth weight on health outcomes is asymmetric. One might assume that the newborns with very low birth weight are exposed at most. Therefore, consider now new variable:

$$bw_i = \begin{cases} \text{very low} & \text{if} \quad bwght_i < 2000 \\ \text{low} & \text{if} \quad 2000 \geq bwght_i \geq 2500 \\ \text{normal} & \text{if} \quad bwght_i > 2500 \end{cases} \tag{4}$$

Estimate the multinomial logit for the above defined dependent variable. And compare estimates with the logit regression from the first part of exercise. Discuss differences.

(vii) Using relative risk ratio interpret the estimates of interest.

(viii) Run now the standard linear regression that explains the logged birth weight and use the same set of explanatory variables as in previous points. Apply the robust standard errors. Compare the significance of estimates between linear regression and logit/multinomial logit. Try to explain differences.