Geneva Graduate Institute (IHEID)

Econometrics II (EI062), Spring 2025

Marko Mlikota

Problem Set 2

Due: Sunday, 23 March, 23:59

- Prepare concise answers.
- State clearly any additional assumptions, if needed.
- You are encouraged to collaborate in groups but the final write-up should be individual.
- Submit your solutions, along with any code (if applicable), in a **single pdf file** through **Moodle**. If you choose to write your solutions by hand, please make sure your scanned answers are legible.
- Grading scale:

5.5	default grade
6	absolutely no mistakes and particularly appealing write-up (clear and concise answers, decent formatting, etc.)
5	more than a few mistakes, or single mistake and particularly long, wordy answers
4	numerous mistakes, or clear lack of effort (e.g. parts not solved or not really attempted)
1	no submission by due date

Problem 1

This problem is based on Griliches and Mairesse (1995, NBER Working Paper 5067, "Production Functions: The Search for Identification"). You can download the data from Moodle either in ASCII format GMdata.RAW or in Stata format GMdata.dta. There are nine variables: *index* (firm ID), *sic3* (3 digit SIC), *yr* (year $\in \{73, 78, 83, 88\}$), *ldsal* (log of deflated sales), *lemp* (log of employment), *ldnpt* (log of deflated capital), *ldrnd* (log of deflated R&D), *ldinv* (log of deflated investment). Consider the model

$$ldsal_{it} = \alpha_i + \beta_1 lemp_{it} + \beta_2 ldnpt_{it} + u_{it}. \tag{1}$$

(a) Compute cross-sectional summary statistics (by year) for the following variables: *ldsal*, *lemp*, *ldnpt*. For each year and each variable report: mean, median, standard deviation, minimum, maximum, 5th percentile, 95th percentile. Generate box plots, one for each variable, placing the years next to each other. Do you see a time trend?

(b) Now, let's create a balanced panel and eliminate firms for which you don't have observations for all four years. How many firms do you loose?

(c) Compute the Random Effects (RE) estimator of $\beta = (\beta_1, \beta_2)'$, i.e., estimate a pooled OLS regression of $ldsal_{it}$ on $lemp_{it}$ and $ldnpt_{it}$ along with an intercept, putting α_i into the error term. State the assumptions needed for consistency of the RE estimator. Are they likely to hold?

(d) Compute the Fixed Effects (FE) Within (FE-W) estimator of β. State the assumptions needed for its consistency. Are they likely to hold?

(e) Compute the FE First Difference (FE-FD) estimator of β. State the assumptions needed for its consistency. Are they likely to hold?

(f) Derive the asymptotic distribution of the FE-W estimator.

(g) Compute the standard errors (i.e. estimates of the standard deviations) of your FE-W estimates of β_1 and β_2. You can base your calculations on the asymptotic variance you derived in the previous exercise, or you can use a command from a software package as long as you can make sure it is based on an appropriate formula.[1]

(h) Compute the standard errors for your FE-W estimates also based on clustered bootstrapping. This is analogous to classical bootstrapping, but, to get a valid panel dataset, you only draw cross-sectional units (firms) with replacement, and for the drawn firms you take all the time

[1] Pay attention not to use any simpler formula that assumes homoskedasticity and/or no serial correlation. In addition, built-in commands in most statistical packages may have finite-sample degree-of-freedom-adjustment terms (in the style of dividing by $(n-1)$ instead of n in the sample variance). While these do not matter asymptotically, in finite sample you may see differences. If you would like to code by hand such finite-sample adjustment terms, see Chapter 17.12 of the Hansen textbook.

periods. Set the number of bootstrap samples $B = 1000$ and take a sample size of n (your actual sample size).
Hint: To facilitate your coding process, first consider a single bootstrap sample, then verify your code works for $B = 10$ or $B = 100$, and only once you solved the whole problem set, take $B = 1000$, as it might take a long time to execute.

(i) Now, instead of creating a balanced panel as in (b), use the full data set (an unbalanced panel) to re-compute the FE-W estimator in (d).

(j) Compute the standard error estimates for your FE-W estimates in (g) based on clustered bootstrap.