Project 1:
Linear Panel Data and Production Technology

Due: Sunday October 5th, 2025, at 22:00

1 Cobb-Douglas Production

In economic models firms are often assumed to have constant returns to scale in production. If the inputs in production are capital (K) and labor (L), then *constant returns to scale* means that the production function (F) satisfies

$$F(\lambda K, \lambda L) = \lambda F(K, L)$$

for all inputs (K, L) and (scales) $\lambda \geqslant 0$. In this assignment we consider a *Cobb-Douglas* production technology, meaning that production technology takes the form

$$F(K, L) = AK^{\beta_K} L^{\beta_L},$$

with $A(> 0)$ representing *total factor productivity* (TFP), and the pair (β_K, β_L) parameters. Note that, while K and L are (at least in principle) observable, we cannot hope to measure *all* factors determining productivity. Hence, A is inherently unobservable in that it aggregates both time-varying and time-invariant unobservable productivity factors, such as managerial talent, power outages, labor strikes, and supply-chain bottlenecks, to name a few.

2 Data

In order to empirically investigate the production technology of firms (i), you are provided with *panel data* for *French manufacturing firms*. Specifically, the dataset covers $N = 441$ firms in the period 1968–79. The accompanying file, `firms.csv`, contain respectively the variables `ldsa` (log of deflated sales, y_{it}), (log of employment, ℓ_{it}), and `lcap` (log of adjusted capital stock, k_{it}).[1] We will treat deflated sales as the output of production. Adopting a Cobb-Douglas form for production thus leads to the relationship

$$y_{it} = \beta_K k_{it} + \beta_L \ell_{it} + v_{it},$$

[1] Each year, the cross-sectional means have been subtracted from the original data. We will ignore this (minor) issue throughout.

where $v_{it} := \ln A_{it}$ represents the logarithm of TFP.

The data is provided in the CSV-file, `firms.csv`, which can be read using the pandas (imported as `pd`) command, `dat = pd.read_csv('firms.csv')`. Then individual variables are in the columns of the dataset, and are extracted as numpy arrays (e.g., `dat.ldsa.values`).

3 Assignment

Formulate an econometric model capturing Cobb-Douglas production. Using all years of data provided, estimate the model parameters (β_K, β_L), and test the hypothesis that production exhibits constant returns to scale. You are allowed to use any of the linear unobserved effects models for panel data discussed in the course and their associated estimators. Discuss the assumptions required for your test to be valid.

4 Hints

(1) Remember to properly define all variables and symbols employed and distinguish between them. For example, you should distinguish between the true parameter and an estimate thereof. Strive to employ the notation used in the course/textbook. Make use of boldface and capitalization to avoid confusing scalars, vectors and matrices. Specify dimensions whenever confusion may arise.

(2) When using an estimation procedure, carefully discuss the assumptions required to derive the estimator and establish properties thereof. Assess whether these assumptions are likely to be satisfied in the current empirical setting. Don't just copy the math; relate to the current setting. If not, what are the consequences for the estimator in question (and your results)? Strive to provide a real-world example of behavior that might invalidate a given assumption, carefully linking the behavior or mechanisms to the mathematical symbols in the model.

(3) If you rationalize several model specifications and associated estimates, discuss which one seems the most appropriate and justify your decision (e.g., based on formal statistical testing or economic reason).

(4) Be precise about the statistical tests you use for testing various hypotheses. Explain which null hypothesis you are testing and the alternative you are testing against, how the test statistic is constructed, the decision rule you employ, and the conclusion you reach. If a variance (matrix) has been estimated, discuss the assumptions invoked for consistency. If several choices are possible, justify your choice.

5 Formal Requirements

- You must hand in a report that presents the econometric model, presents your estimation results and results of formal statistical tests (including interpretation and statements on economic and statistical significance), and discusses the potential weaknesses of the model, data and approach. If you present many estimates of the same parameters (e.g., estimators based on different assumptions, or varying the controls or sub-sample used), it may be helpful to present the estimates together in one table to facilitate comparison.

- The report must be written in English using an academic language and structure and uploaded to FeedbackFruits via Absalon as a single PDF file.

- The report must be structured as at most five pages of main text (including mathematics) followed by at most two pages of output. Use the following formatting:

  - For the main text (and mathematics), you must use fontsize = 12p, line spacing = 1.5, and page margins = 2.5 cm (as used in this document). Note that this may differ from your idea of a "normal" page.
  - The ouput can be any (relevant) tables, graphs or images as long as they are properly formatted, labelled and readable. Place the output at the end of your report, starting on a new page. Do not worry about this order being inconvenient for the reader.

- Along with your report, you must upload a compressed zip-folder with all the Python code needed to replicate your results. Make sure that your code is transparent and runs with only minor modifications (e.g., changing relevant paths). There is no character limit on the submitted Python code.

- You are allowed (and strongly encouraged) to work in groups of up to three people. List all group members on the front page of your report in alphabetical order of surnames.

- The general assessment criteria are given on the course website in Absalon, where you can also find some tips on how to write and structure your report.