SOK-3011 Linear Models in Economics

Assignment 1

Submission due date: October 15, 23:59

Oral presentation: October 17, 10:15—14:00

Instructions:

This assignment consists of two parts. In part 1, submit a draft (PDF) documenting your theoretical analysis. In part 2, you must collect a dataset and provide a brief analytical discussion. You must submit a written report (including your code) and deliver an oral presentation of your findings using the collected data. For your final submission, you will be analyzing firm-level production and cost data (more details provided during the course). You can work on the same dataset gathered in Part 2 for the final submission.

Part 1: Theoretical analysis

Question 1a: Cobb-Douglas technology

Suppose the true technology is Cobb–Douglas:

$$\ln y = \beta_0 + \alpha_1 \ln x_1 + \alpha_2 \ln x_2 + \varepsilon.$$

- What do we mean by the *output elasticity* ε_i and the *elasticity of scale* ε? In the above specification, determine ε. Classify IRS/CRS/DRS using ε.
- Under competitive profit maximization (or cost minimization), relate *cost share* $s_i = \frac{w_i x_i}{\sum_k w_k x_k}$ to the output elasticity. How could you use observed s_i to check model consistency?

Question 1b: Translog production technology

Consider the Trans-log production form:

$$\ln y = \beta_0 + \sum_i \beta_i \ln x_i + \tfrac{1}{2} \sum_i \sum_j \beta_{ij} \ln x_i \ln x_j.$$

- Explain what the β_{ij} terms allow you to capture (relative to Cobb–Douglas) in terms of *MRTS* and *elasticity of substitution*. If you *mean-center* logs , *i.e.*, replace $\ln x_i$ by $lnx_i - \overline{\ln x_i}$, what interpretation do β_i have at the *sample mean*?
- Describe two practical estimation issues with Translog forms and a simple way to handle each.

Part 2: Data analysis (Submit the output as a report and make oral presentation)

Question 2a: Data collection

Collect a firm-level dataset that includes at least the following variables: Output–The quantity of goods produced; Inputs–At least two types of inputs; Costs–Total production cost or variable cost, or the input prices. Data on the output price or revenue will also be useful. Ensure that the dataset contains a sufficient number of observations (no. of firms > 25) for meaningful analysis.

- Provide a brief description of the dataset (source, number of firms, number of variables, etc.). Provide some descriptive statistics of the dataset (mean, standard deviation, min, max for each variable).

Note: If you are unable to collect raw data, you may use datasets available in R packages, such as *sfaR*, *micEcon*, *rDEA*, *deaR*, *Benchmarking*, or others. A good resource to find relevant datasets is (https://vincentarelbundock.github.io/Rdatasets/datasets.html). Even if using an existing dataset, you must describe it properly, citing the original sources and explaining its purpose, industry, and time period.

Question 2b: Estimation of a production function

- Estimate one production function (possibly a linear one) using your data. Interpret the estimated coefficients. Analyze and discuss (i) productivity, (ii) returns to scale, and (iii) input substitution, based on the estimated model. Perform a goodness-of-fit analysis. Apply Ramsey's Regression Equation Specification Error Test (RESET) to check for model specification errors.