

Problem Set 2

Due on Thursday, November 6th @1pm

Submit (1) your source code via KLMS,

(2) a written copy of your work & printed source code in class

In this problem set, you will replicate some of the results in Nerlove (1963). In his paper, Nerlove analyzed a cost function for 145 American electricity companies. This work is discussed in Section 1.7 of Hayashi and the empirical exercise in Chapter 1 of Hayashi. Please read Section 1.7 of Hayashi before you work on the following questions

There are optional questions¹ in this problem set. Although they are optional, I highly recommend that you try them both for a better learning experience and a better grade.

Please program in Matlab², R, or Python. **Please do NOT use statistical software packages such as STATA, SAS or SPSS or statistical analysis libraries in R or Python - I want you to program matrix operations directly to implement OLS.**

You can find the data file and template Matlab programming files (where you can add your own code) in our Google Drive:

- nerlove.asc: the data file, you may read this data in your source code

This text file has data on 145 electric utility companies in 1955 for the following:

Column 1: total costs (TC) in millions of dollars

Column 2: output (Q) in billions of kilowatt hours

Column 3: price of labor (PL)

Column 4: price of fuels (PF)

Column 5: price of capital (PK)

The data are ordered by level of output. Also NOTE that columns 4 and 5 are in reverse order to that we use in the estimation equations for this problem set.

- nerlove.mat: nerlove.asc in a Matlab data format

¹If you get the perfect score on optional questions, you will get additional points of $\sigma/5$ toward your total numeric grade (out of 100) of this course, where σ is a standard deviation of total numeric grades at the end of the semester. This optional credit will be used only towards moving one letter grade up. That is, I will first assign letter grades to your total numeric grades before the optional credits are added, and if “your total numeric grades+optional credits” $\geq \min\{x|x\text{ is a total numeric grade (before optional credits) that received one letter grade higher than yours}\}$, your final letter grade will go up by one level. In the past, about 10% of the students benefitted from this.

²Matlab is available for you to download from kftp.kaist.ac.kr

- main.m: This matlab file reads ‘nerlove.mat’, and you can start programming by adding your code in this file. You may run this program first to see what it does.
- ols.m: This Matlab file implements the basic OLS function in Matlab. You can call this function in your main.m. The current version calculates the OLS coefficients only, so you may want to add your code to calculate other statistics such as standard errors of your OLS coefficients and R^2 .

And please make sure that every file is in the same folder.

1. Nerlove starts from the assumption of the following production function:

$$Q_i = A_i(Labor_i)^{\alpha_1}(Capital_i)^{\alpha_2}(Fuel_i)^{\alpha_3}$$

(a) What does $\alpha_1 + \alpha_2 + \alpha_3$ mean?

(b) (*Optional, Difficult - you need to study the producer theory in microeconomics first*) This production function implies the following equation on the total cost:

$$\log(TC_i) = \frac{1}{r} \log(Q_i) + \frac{\alpha_1}{r} \log(PL_i) + \frac{\alpha_2}{r} \log(PK_i) + \frac{\alpha_3}{r} \log(PF_i) + \log \left[r(A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/r} \right], \quad (1)$$

where $r \equiv \alpha_1 + \alpha_2 + \alpha_3$. Derive this equation (1). (Hint: Set up a cost minimization problem of firm i and plug in the optimal factor demand back to the total cost.)

2. The equation (1) can be estimated by the OLS estimation of the following unrestricted Cobb-Douglas specification:

$$\log(TC_i) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log(PL_i) + \beta_4 \log(PK_i) + \beta_5 \log(PF_i) + \epsilon_i \quad (2)$$

Report the OLS coefficients, their standard errors, t-test on the significance of each β , and R^2 . Comment on your results.

3. You can restrict the previous model (2) by directly imposing the hypothesis $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$, which results in the following restricted model:

$$\log \left(\frac{TC_i}{PF_i} \right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log \left(\frac{PL_i}{PF_i} \right) + \beta_4 \log \left(\frac{PK_i}{PF_i} \right) + \epsilon_i. \quad (3)$$

Run this regression using the data. Report the OLS coefficients, their standard errors, and R^2 . Comment on or interpret your results.

4. Report your F-test result on the hypothesis $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$ for the estimation model (2) (Hint: Use this F-test formula that uses the sum of squared errors of the unrestricted model (SSR_U) and restricted model (SSR_R), $F\text{-ratio} = \frac{(SSR_R - SSR_U)/\#r}{SSR_U/(n-K)}$.) What does this hypothesis testing try to find out? Comment on your results.

5. Plot the residuals (y-axis) over each of your independent variables including $\log Q_i$ (x-axis) for the OLS estimation of the model (3). What do you observe?

The next set of questions are adapted from p.77-79 of Hayashi.

As you can observe from the plot in the previous question, the plot of residuals suggests a nonlinear relationship between $\log(TC)$ and $\log(Q)$. Nerlove hypothesized that estimated returns to scale varied with the level of output. Following Nerlove, divide the sample of 145 firms into five subsamples or groups, each having 29 firms. (Recall that since the data are ordered by level of output, the first 29 observations will have the smallest output levels, whereas the last 29 observations will have the largest output levels.) Consider the following three generalizations of the restricted model in Question 3.

Model 1: Both the coefficients and the error variance differ across groups.

Model 2: The coefficients are different, but the error variance is the same across groups.

6. For Model 1, the coefficients and error variances specific to groups can be estimated from

$$y^{(j)} = X^{(j)}\beta^{(j)} + \epsilon^{(j)}, (j = 1, \dots, 5),$$

where $y^{(j)}$ (29×1) is the vector of the values of the dependent variable for group j , $X^{(j)}$ (29×4) is the matrix of the values of the four regressors for group j , $\beta^{(j)}$ (4×1) is the coefficient vector for group j , and $\epsilon^{(j)}$ (29×1) is the error vector. The second column of $X^{(5)}$, for example, is $\log(Q_i)$ for $i = 117, \dots, 145$. Model 1 assumes conditional homoskedasticity $\epsilon^{(j)}\epsilon^{(j)'}|X^{(j)} = \sigma_j^2 I_{29}$ within (but not necessarily across) groups.

Estimate Model 1 by OLS. Comment on the returns to scale in each of the five groups. What is the general pattern of estimated scale economies as the level of output increases? What is the general pattern of the estimated error variance as output increases?

7. You can estimate Model 2 using the following dummy variables.

$$\log\left(\frac{TC_i}{PF_i}\right) = \sum_{j=1}^5 D_{j,i} \left\{ \beta_1^{(j)} + \beta_2^{(j)} \log(Q_i) + \beta_3^{(j)} \log\left(\frac{PL_i}{PF_i}\right) + \beta_4^{(j)} \log\left(\frac{PK_i}{PF_i}\right) \right\} + \epsilon_i,$$

where the j -th dummy variable for the firm i is defined as

$$D_{j,i} = 1 \quad \text{if firm } i \text{ belongs to the } j\text{-th group,} \quad 0 \quad \text{otherwise.}$$

Estimate this model by OLS. Verify that the OLS coefficient estimates here are the same as those in the previous question.

8. (Optional) Prove that the OLS coefficients and SSR (sum of squared residuals) are the same for both Model 1 and Model 2.