Itroductory Econometrics for Business and Economics

September/October 2021

Enzo D'Innocenzo

Assignment

The assignment grade counts for 15% of the final grade for this course.

Some important remarks:

- **Names** and **student numbers of all group members** should be on the front page of the solutions.
- Name the document in the following format:

  LastName1_LastName2_LastName3_LastName4.pdf.

- You should be able to use LaTeX, which is highly recommended. Otherwise, you may also use Word or any other text editor but please hand in a **.pdf** document with your solution presented in a formal fashion.
- It is important to present your results logically and succinctly while making sure that you have included enough information to prove that you have researched and studied the matter thoroughly.
- I strongly recommend to use `R` for solving the assignment. You can also use MATLAB or PYTHON. **No** Eviews, Stata, or anything else.
- Present your codes, properly commented, at the end of the pdf, and upload them with the assignment pdf on Canvas. If using LaTeX, I recommend the package `listings`.

In this assignment you are asked to model the dataset used by Marc Nerlove in the influential studies on returns to scale in the United States electricity generation market. Download the file `NervloveData63.RData` and open it with `R`. This file contains the following set of variables:

1. *TC (totcost)*: Total cost in 1970 MM USD,
2. *Q (output)*: Output billions of kilowatt hours,
3. *PL (plabor)*: Price of labor,
4. *PF (pfuel)*: Price of fuel,
5. *PK (pkap)*: Price of capital.

Question 1

Estimate by OLS the following log-log Multiple Regression model:

$$\log(TC_i) = \beta_0 + \beta_1 \log(Q_i) + \beta_2 \log(PL_i) + \beta_3 \log(PF_i) + \beta_4 \log(PK_i) + u_i,$$

for $i = 1, \ldots, 145$. Interpret the estimated coefficients in the log-log equation. Are the OLS coefficient unbiased? What they do represent in practice? Test the coefficients and comment the obtained results.

Question 2

Check the correlations between all pairs of X variables and give some comments. If the there is reason to expect that the regressors are perfectly multicollinear, what does multicollinearity do to your regression? Will the overall fit be affected? Test again the coefficients by using heteroskedasticity-robust standard errors. Test also the homogeneity restriction $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$. In particular, write the hypothesis in the form $\boldsymbol{R\beta} = \boldsymbol{r}$ and calculate the F-statistics, find the critical value and discuss your conclusions.

Question 3

Extrapolate and plot the residuals against the fitted values. Also, depict a QQ-plot. What can we learn from these two graphs? Are the OLS assumptions satisfied? In particular, verify the following assumptions: Strict exogeneity, spherical error variance, and the fourth moments conditions.

Question 4

Plot the residuals against $\log(Q)$. Does the plot suggests a nonlinear relationship between the two variables? Discuss the plot. One of the LS assumption is clearly violated, which one? How can you capture this nonlinear relationship in a multiple regression model?

Question 5

Estimate by OLS a new nonlinear Multiple Regression model:

$$\log(TC_i) = \beta_0 + \beta_1 \log(Q_i) + \beta_2[\log(Q_i)]^2 + \beta_3 \log(PL_i) + \beta_4 \log(PF_i) + \beta_5 \log(PK_i) + u_i,$$

for $i = 1, \ldots, 145$. How can we interpret regression results from a squared logarithm specification? Plot again the residuals against the fitted values and compare it with the same plot asked in **Question 3**.

Question 6

Divide the sample of 145 firms into five subsamples, each having 29 firms. The first 29 observations will have the smallest Q levels, whereas the last 29 observations will have the largest output levels 29 firms. Report in a table descriptive statistics for each subsamples and comment the results.

Question 7

Estimate a regression model for each subsample by OLS. This model assumes conditional homoskedasticity within groups, but not across groups. What is the general pattern of the estimated error variance as $\log(Q)$ increases?

Question 8

Estimate a regression model by OLS which have different coefficient across groups, but same error variance. To estimate this model, define a set of dummy variables for each group, such that

$$D_{ij} = \begin{cases} 1 & \text{if the firm i belongs to the j-th group,} \\ 0 & \text{otherwise.} \end{cases}$$

Are the OLS coefficient estimates the same as those in **Question 7**? Why?

Question 9

Estimate a restricted version of the regression model in **Question 8**. The new model, has different coefficient across groups only for the intercepts and the variable $\log(Q)$. The other variables have the same structure of the model estimated in **Question 1**.

Question 10

Using the model selection techniques you know, propose a model for $\log(TC)$. Feel free to consider polynomial terms, interaction variables and logarithmic terms into your model. Compare the candidate versions of your regression and give reasons why you consider your final model to be the best.