Financial Econometrics Workshop 1

Binzhi Chen

July 16, 2020

1 Introduction

In this workshop we will introduce some basic statistics and graph tools in STATA. Most of descriptive statistics would be commonly use in data analysis. Linear regressions and constraint linear regressions will also be introduced in this workshop.

2 Input Data

The first step in STATA is input data, make sure your .dta file is in the right working directory. In this workshop we will use the dataset from *Nerlove(1963)*.

Menu:

File → Change Working Directory → STATA Workshop

Command:

```
use nerlove63.dta,clear #open the data file
label data "Returns to scale in electricity supply, Nerlove, 1963" # set labels for the whole dataset
help describe
describe
```

> **Note:**
>
> You may use `help` command to check the meaning of the code.

> **Note:**
>
> For simplicity, you may use `d` to replace *describe* in STATA.

3 Descriptive Analysis

In this section, you will learn how to type descriptive command to show the descriptive analysis result.

Command:

```
list totcost output
list totcost output in 1/5 #select the head of the dataset
summarize totcost #decriptive analysis or you can use su instead
summarize totcost,detail #more details of descriptive analysis
return list
display "The coefficient of variation is" r(sd) / r(mean)
pwcorr totcost output plabor,sig star(.05) #pairwise correlation put stars for the correlation at 95% significance level
```

Note:

`summarize [variable],detail` command has more details, including skewness and kurtosis.

Note:

All results are saved in `r()` command while you `return list`.

4 Graph

We may use some graph to visualize the data.

Command:

```
histogram totcost,frequency #show the graph for the variable. y-axis is the frequency
scatter totcost output #scatter plot
scatter totcost output,mcolor(maroon)
scatter totcost output,msymbol(sh)
generate n =_n #_n stands for observation n
scatter totcost output,mlabel(n) mlabpo(6) #mlabpo makes all labels in six o'clock
twoway (sc totcost output) (lfit totcost output)
```

Note:

Options in `scatter [variable],options` command leads to different types of graph.

5 Data Transformation and Linear Regression

Since the original equation is a Cobb-Douglas function:

$$TC_i = \delta_i Q_i^{1/r} (P_{L,i})^{\alpha_1/r} (P_{K,i})^{\alpha_2/r} (P_{F,i})^{\alpha_3/r}$$

you may take logs to get the linear equation:

$$lnTC_i = \beta_1 + \frac{1}{r} lnQ_i + \frac{\alpha_1}{r} lnP_{L,i} + \frac{\alpha_2}{r} lnP_{K,i} + \frac{\alpha_3}{r} lnP_{F,i} + \epsilon_i$$

Command:

```
generate lntc = log(totcost)
generate lnq  = log(output)
generate lnpl = log(plabor)
generate lnpf = log(pfuel)
generate lnpk = log(pkap)
reg lntc lnq lnpl lnpf lnpk #linear regression
vce #Variance covariance matrix estimation
predict fitted_value
predict e1,residual
test (lnq = 1)(lnpl + lnpk + lnpf = 1) #joint hypothesis
```

Note:

Where `fittedvalue` is the fitted value of dependent variable $\hat{y}$.

Note:

`e1` is the fitted value of residuals $\hat{\epsilon}_i$

6 Restricted Linear Regression

If we manually put restrictions on variables. We may use RLS:

Command:

```
constraint def 1 lnpl+lnpf+lnpk=1 #define constraint 1
constraint def 2 lnq=1 #define constraint 2
cnsreg lntc lnq lnpl lnpf lnpk,c(1-2)
```

Note:

Where the sum of independent variables $lnpl$, $lnpf$, $lnpk$ is 1. lnq equals to 1.