

Efficiency analysis using STATA

Kwok Tong Soo

April 2013.

Contents:

1. Introduction to STATA.
2. Data Envelopment Analysis in STATA.
3. Stochastic Frontier Analysis in STATA.

1. Introduction to STATA

STATA is a powerful and flexible software. You should already have some experience with using STATA from the Econ420 sessions. In this documentation, STATA commands are in **Courier New font**.

STATA has extensive documentation accessible from the programme. Type **help regress** to call up the help file for the regress command. Then in the help file choose Also See → regress (PDF) for the PDF manual associated with regress.

STATA is also connected to the internet. It performs regular updates, and allows users to download user-written packages which extend the commands available. Type **ssc hot** for a list of the most popular downloads. Type **net search X** to perform a search for topic X.

1.1. Data organisation in STATA

STATA has a powerful function to help organise your data, the **reshape** command. However, it is quite difficult to use. My recommendation is that you organise your data in the appropriate format in Excel, then enter it into STATA using the **import excel** command.

Your data should be in long format. This means that each firm-year has one row in Excel, with variables in columns. For example:

| Firm | Year | Var1 | Var2 | Var3 |
|------|------|------|------|------|
| X | 1 | | | |
| X | 2 | | | |
| Y | 1 | | | |
| Y | 2 | | | |
| Z | 1 | | | |
| Z | 2 | | | |

STATA also has **merge** and **append** commands. **merge** is used when you want to add more variables with the same observations to your dataset. **append** is used when you want to add more observations with the same variables to your dataset. These commands are quite easy to use and we will demonstrate their use in the session.

In addition, STATA has powerful variable generating commands, **generate** and **egen** (extensions to generate). So all you need is to enter the basic data into STATA and use these commands to generate the variables that you need for your analysis.

1.2. Using Do-files

One of the most powerful features of STATA is the facility for saving your commands in Do-files. This allows you to easily save and replicate your work, and facilitates collaborations. More advanced users can learn to programme in STATA, starting with writing loops, and then making use of the matrix programming facilities in STATA.

2. Data Envelopment Analysis (DEA) in STATA

The main idea behind DEA is to first determine the decision-making-units (DMU) which are on the frontier, and then calculate the efficiency of other DMUs relative to this frontier. DMUs use inputs in order to produce output, and may be thought of as either choosing inputs to minimise cost for a given output level (input orientation), or choosing output to maximise revenue for a given input level (output orientation). The production function may have constant returns to scale, variable returns to scale, or non-increasing returns to scale.

STATA does not have a built-in command to perform DEA. However, there is a user-written command, **dea**, which you can download and use. To do this, type **net install st0193**. Type **help dea** to call up the help file. The syntax is:

```
dea ivars = ovars [if] [in] [, options]
```

Where the options are:

| | |
|------------------------------------|---|
| rts(crs vrs drs nirs) | specifies the returns to scale. The default is rts(crs) |
| ort(in out) | specifies the orientation. The default is ort(in) |
| stage(1 2) | specifies the way to identify all efficiency slacks. The default is stage(2) |
| trace | save all sequences and results from Results window to dea.log |
| saving(filename) | save results to filename. |

The example do-file **DEA.do** provides an example of using the **dea** command. It seeks to replicate some of the results in Chen and Soo (2010) (available from <http://www.accessecon.com/Pubs/EB/2010/Volume30/EB-10-V30-I4-P249.pdf>), and provide additional results for a second-stage regression. In addition, it provides an insight into some additional useful commands, and how your data should be structured. Further examples of DEA can be found in Ji and Lee (2010).

3. Stochastic Frontier Analysis (SFA) in STATA

STATA has a built-in command to estimate SFA. This is essentially a regression analysis where the error term consists of a random error and an inefficiency term, and again can be estimated for both production and cost functions. The syntax of the **frontier** command is:

```
frontier depvar [indepvars] [if] [in] [weight] [, options]
```

Type **help frontier** to see the options available. The most important options are:

| | |
|-------------------------------|--|
| distribution(distname) | specifies the distribution for the inefficiency term as half-normal (hnormal), exponential , or truncated-normal (tnormal). The default is hnormal . |
| cm(varlist) | fit conditional mean model; may be used only with distribution(tnormal) . |
| uhet(varlist) | explanatory variables for technical inefficiency variance function |
| vhet(varlist) | explanatory variables for idiosyncratic error variance function |
| cost | fit cost frontier model; default is production frontier model |
| vce(vcetype) | vcetype may be oim , opg , bootstrap , or jackknife |

The examples in this section are taken from the STATA manual. Note that it is also possible to perform SFA for panel data: see **help xtfreontier** for more details.

References

Chen, Ching-Fu and Kwok Tong Soo (2010), "Some university students are more equal than others: Efficiency evidence from England", *Economics Bulletin* 30(4): 2697-2708.

Ji, Yong-Bae and Choonjoo Lee (2010), "Data envelopment analysis", *Stata Journal* 10(2): 267-280.