The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–8

Free Disposal Hull(FDH) Analysis for Efficiency Measurement: An update to dea

Byoungin Lim
Nextree Soft Co., Ltd.
Seoul/Republic of Korea
byleem@nextree.co.kr

Kyoungrok Lee
Korea National Defense University
Seoul/Republic of Korea
rock6371@naver.com

Choonjoo Lee
Korea National Defense University
Seoul/Republic of Korea
bloom.rampike@gmail.com

Abstract. This paper introduces an author written Free Disposal Hull(FDH) model for efficiency measurement in Stata. The FDH model was introduced and developed from the discussions of (Deprins, Simar, and Tulkens 1984),Thrall (1999), Cherchye and Post (2000) in the field of production economics. Free Disposal Hull model assumes the free disposability relaxing the convexity assumption in defining the production possibility set from the observations. So, FDH is considered as an alternative approach to DEA for efficiency measurement. This update was required to improve the computation time of the original `dea` program and reply to the requesters on the FDH option in `dea`. It shows how to use FDH option in `dea` command.

Keywords: st0001, Free Disposal Hull(FDH), efficiency, production possibility set, convexity, free disposability.

1 Introduction to FDH model

Free Disposal Hull(FDH) model is a non-parametric method to measure the efficiency of production units or Decision Making Units in other words.

The FDH model was conceptualized, formulated, developed (Deprins, Simar, and Tulkens 1984), and extended (Lovell et al. 1994). And as such that the basic DEA models in Stata (Ji and Lee 2010) triggered off the outlier detection and the statistical inference issues in Stata (Tauchmann 2012), FDH model option in `dea` will provide Stata users with more program options to compare the results among those programs in Stata since they have their own distinct characteristics.

FDH model relaxes the convexity assumption of basic DEA models. The computational technique to solve FDH program considers the mixed integer programming problem compared to the DEA model with a linear programming problem.

The main purpose of this paper is to provide the updated `dea` and to explain about the FDH efficiency analysis using it in Stata. The next section describes the FDH model concepts and calculations comparing with basic DEA models (Ji and Lee 2010) in Stata.

 st0001

And then this paper illustrates the features and options of the FDH program.

2 The Basics of FDH Model

If specific inputs can produce specific outputs, the pairs of these inputs and outputs are producible, so the pairs are called the production possibility set.

In Figure 1, if any activity (x, y) belongs to the Production Possibility Set(P), then the activity (tx, ty) belongs to P for any positive scalar t. This property is called constant returns-to-scale(CRS) assumption. Charnes, Cooper, and Rhodes (1978) developed the CCR(Charnes-Cooper-Rhdes) model to evaluate the production units' efficiency assumming CRS. This assumption can be modified to allow the production possibility set with different postulates. BCC(Banker-Charnes-Cooper) model is representative using by variable returns-to-scale(VRS). It is characterized with increasing returns-to-scale(IRS), decreasing returns-to-scale(DRS), and constant returns-to-scale(CRS).

The production possibility set of FDH model is obtained by defining it differently with CCR and BCC models. In the CCR and BCC models, if (x1, y1) and (x2, y2) belong to the production possibility set, then the (a(x1+x2), b(y1+y2)) with any positive scalar a, b is also considered to be in the same production possibility set. The axiom is called convexity.

Free disposability means if a specific pair of input and output is producible, any pairs of more input and less output for the specific one are also producible. FDH model allows the free disposability to construct the production possiblity set. Accordingly, the frontier line for FDH model is developed from the observed inputs and outputs allowing the free disposability.

In Figure 2, the shape of production possibility set in FDH is stepwise. The frontiers determined for FDH model are presented considering two inputs and one output for 6 production units labeled A through F. In BCC model, DMUs A, B and C are efficient but A, B, C and F are efficient in FDH model. The efficiency of observation E in BCC model is defined as $\theta_{E,BCC,input} = \overline{OE_2}/\overline{OE}$. But the efficiency of observation E in FDH model is defined as $\theta_{E,FDH,input} = \overline{OE_1}/\overline{OE}$.

One version of FDH model aims to minimize inputs while satisfying at least the given output levels. This is called the input-oriented model. The other one is called the output-oriented model that attempts to maximize outputs without requiring more inputs. The scores of efficiency in FDH model are between 0 and 1. And under input-oriented condition, the efficiency scores of FDH input-oriented model are always greater than the ones ofinput-oriented variable returns to scale(VRS) model. Also, the efficiency scores of input-oriented VRS model are always greater than those of input-oriented CRS model. In other words, the production possibility set of FDH model is subset of VRS model as well as CRS model.

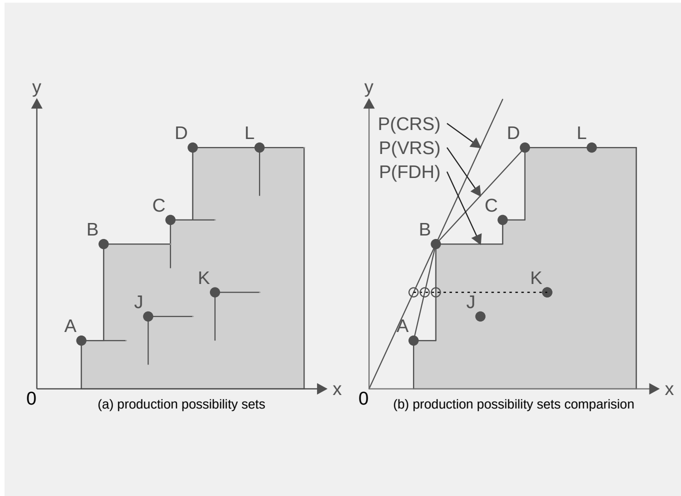

Figure 1: production possibility sets in CRS, VRS, and FDH

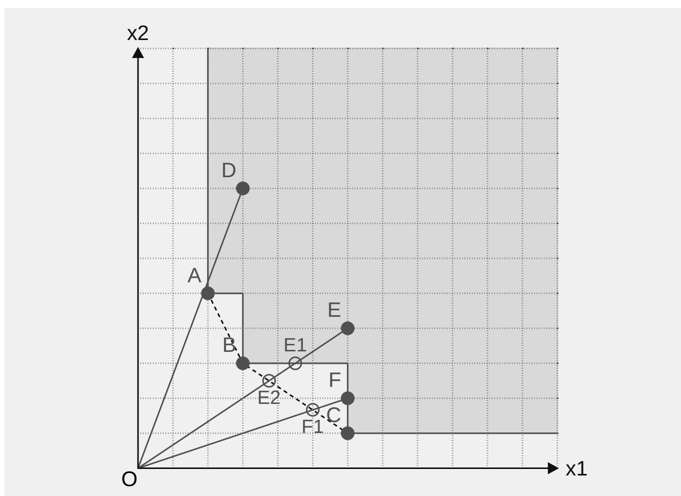

Figure 2: Input Oriented FDH Efficiency Measures

$$\theta_{FDH,input} \geq \theta_{VRS,input} \geq \theta_{CRS,input} \tag{1}$$

Under the CRS assumption and input-oriented condition, the formulation for DMU$_k$ is

$$\begin{aligned} &\min \theta_k \\ \text{Subject to } &\theta x_k - X\lambda \geq 0 \\ &Y\lambda \geq y_k \\ &\lambda \geq 0 \end{aligned} \tag{2}$$

Where λ is a semi-positive vector in R^{n}. In FDH model, two constraints, $e\lambda = 1, \lambda_j \in \{0, 1\}$, are added. $\lambda \in \{0, 1\}$ means that the components of λ are constrained to be bivalent. That is, they must all have values of zero or unity so that together with the condition $e\lambda = 1$ one and only one of the performances actually observed can be chosen.
Accordingly the formulation of FDH model is

$$\begin{aligned} &\min \theta_k \\ \text{Subject to } &\theta x_k - X\lambda \geq 0 \\ &Y\lambda \geq y_k \\ &e\lambda = 1, \lambda_j \in \{0, 1\} \end{aligned} \tag{3}$$

In FDH model, there is a possibility that efficient DMUs have positive slacks to be disposed of. The two-stage method is used to treat the slacks in Equation (4) .
Accordingly, Equation (4) needs to be solved followed by Equation (3).

$$\begin{aligned} &\max \Sigma s^+ + s^- \\ \text{Subject to } &\theta x_k - X\lambda \geq 0 \\ &Y\lambda \geq y_k \\ &e\lambda = 1, \lambda_j \in \{0, 1\} \end{aligned} \tag{4}$$

3 The `dea` command for FDH analysis

3.1 Syntax

The syntax of the `dea` command for FDH analysis is:

`dea` *inputvars(min=1) = outputvars(min=1)* [*if*] [*in*] [`using/` *filename*]`,fdh` [
`ort(i|in|input|o|out|output) stage(integer 2) tol1(real 1e-14)`
`tol2(real 1e-8) trace saving(`*filename*`) replace`]

Where `inputvars` and `outputvars` mean input and output variable lists, respectively.

3.2 Description

`dea` for FDH analysis requires the user to select the input and output variables from the user designated data file and solves FDH models by options specified.

The `dea` program requires initial data set that contains the input and output variables for observed production units. Variable names must be identified by `inputvars` for input variable and by `outputvars` for output variables to allow that `dea` program can identify and handle the multiple input-output data set. And the variable name of DMUs must be specified by “dmu” and the option must be specified by ”fdh” to solve FDH model using `dea` command.

The program has the ability to accommodate unlimited number of inputs/outputs with unlimited number of DMUs. The only limitation is the memory of computer used to run `dea` and the number of observations(DMUs) than the combined number of inputs and outputs to solve the MILP problem. The result file reports the information including reference points and slacks in FDH models. These informations can be used to analyze the inefficient production units, for examples, where the source of inefficiency comes from and how could improve an inefficient unit to the desired level.

`saving`(*filename*) option creates a *filename*.dta file that contains the results of FDH including the information of the DMUs, inputs and outputs data used, ranks of Decision Making Units(DMUs), efficiency scores, reference sets, and slacks. The log file “fdh.log” will be created in the working directory.

The `dea` program requires the input and output variables and data sets and the options to be defined for the FDH model selection. Based on the data and options specified in the `dea` program, the `dea` program conducts the matrix operations and linear programming to produce the results data sets that are available for print or can be used for further analysis.

3.3 Options

`ort(i|in|input|o|out|output)` specifies the orientation. The default is `ort(i)` or `ort(in)`, meaning the input oriented FDH. `ort(o)` or `ort(out)` means the output oriented FDH.

`stage(integer 2)` specifies the way to identify all efficiency slacks. The default is `stage(2)`, meaning the two-stage FDH. `stage(1)` means the single-stage FDH.

`tol1(real 1e-14)` specifies the tolerance of entering or leaving value. see [M-1] **tolerance**

`tol2(real 1e-8)` specifies the tolerance of B inverse matrix value. see [M-1] **tolerance**

`trace` lets all the sequences displayed in the result window and also saved in the “fdh.log” file. The default is to save the final results in the “fdh.log” file.

`saving(`*filename*`)` specifies that the results be saved in *filename*`.dta`. If the same *filename* already exists and `replace` option not specified, the existing *filename* will be replaced by the name of *filename*`_bak_DMYhms.dta`.

`replace` specifies whether or not to replace the log file. if the saving file exists and replace option is not specified, the backup file will be created.

3.4 Saved Results

`dea` for FDH analysis saves the following in `r()`:

Matrices

`r(dearslt)` `n x m` matrix of the results of `dea` command for FDH analysis where `n` is the number of DMUs and `m` is the information depending on the models specified. Rows correspond to the DMUs and columns correspond to the variables including inputs, outputs, rank(of DMUs scores), theta(efficiency scores), ref.(reference DMUs), input and output slacks, and more depending on the models specified.

4 Applications of `dea` command for FDH analysis

4.1 Data

This section provides examples taken using data from Cooper, Seiford, and Tone (2006, 75, Table 3.7) for illustration of the `dea` program. The data of Cooper et al. (2006) consist of five stores that use two inputs, **employees**(number of employees as an input variable) and **area**(the area of floor as an input variable), to produce two outputs, **sales**(the volume of sales as an output variable) and **profits**(the volume of profits as an output variable).

The following shows the data of `cooper_tbl_3_7.dta`.

```
. use cooper_tbl_3_7.dta
. list
```

	dmu	employee	area	sales	profit
1.	A	10	20	70	6
2.	B	15	15	100	3
3.	C	20	30	80	5
4.	D	25	15	100	2
5.	E	12	9	90	8

4.2 Input-oriented two stage FDH model

```
. use cooper_tbl_3_7.dta, replace
. dea  employee area =  sales profit, fdh ort(in) stage(2)
--------------------------------------------------------------------------------

options: ORT(IN) STAGE(2)
FDH-INPUT Oriented DEA Efficiency Results:
                                ref:      ref:      ref:      ref:      ref:
           rank     theta         A         B         C         D         E
dmu:A         1         1         1         0         0         0         0
dmu:B         3        .8         0         0         0         0         1
```

```
dmu:C          5        .6         0         0         0         0         1
dmu:D          4        .6         0         0         0         0         1
dmu:E          1         1         0         0         0         0         1
         islack:   islack:   oslack:   oslack:
       employee      area     sales    profit
dmu:A          0         0         0         0
dmu:B          0         3       -10         5
dmu:C          0         9        10         3
dmu:D          3         0       -10         6
dmu:E          0         0         0         0
-------------------------------------------------------------------------------
. return list
matrices:
            r(dearslt) :  5 x 11
```

4.3 Output-oriented two stage FDH model

```
. use cooper_tbl_3_7.dta, replace
. dea  employee area =  sales profit, fdh ort(out) stage(2)
-------------------------------------------------------------------------------
options: ORT(OUT) STAGE(2)
FDH-OUTPUT Oriented DEA Efficiency Results:
                                ref:      ref:      ref:      ref:      ref:
           rank     theta         A         B         C         D         E
dmu:A          5   1.28571         0         0         0         0         1
dmu:B          1         1         0         1         0         0         0
dmu:C          4     1.125         0         0         0         0         1
dmu:D          1         1         0         0         0         1         0
dmu:E          1         1         0         0         0         0         1
         islack:   islack:   oslack:   oslack:
       employee      area     sales    profit
dmu:A         -2        11         0   .285714
dmu:B          0         0         0         0
dmu:C          8        21         0     2.375
dmu:D          0         0         0         0
dmu:E          0         0         0         0
-------------------------------------------------------------------------------
. return list
matrices:
            r(dearslt) :  5 x 11
```

5 Conclusion

This paper provides the updated `dea` and explains about the option for Free Disposal Hull(FDH) model that measures the productive efficiency without the convexity assumption. FDH model is driven under the assumption of the free-disposability to obtain the production possibility set. The shape of the production frontier looks like stepwise. The FDH efficiency analysis in Stata will allow users to extend their models on non-parametric efficiency measurement as well as stochastic and semi-parametric approaches.

6 References

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* 2: 429–444.

Cherchye, K. T., L., and T. Post. 2000. What Is the Economic Meaning of FDH? A Reply to Thrall. *Journal of Productivity Analysis* 13: 263–7.

Cooper, W., L. M. Seiford, and K. Tone. 2006. *Introduction to Data Envelopment Analysis and Its Use*. 233 Spring Street, New York, NY 10013, USA: Springer Science & Business Media, Inc.

Deprins, D., L. Simar, and H. Tulkens. 1984. Measuring Labor Inefficiency in Post Offices. In *The Performance of Public Enterprises: Concepts and Measurements*, ed. M. Marchand, P. Pestieau, and H. Tulkens, 243–267. Amsterdam: North-Holland.

Ji, Y., and C. Lee. 2010. Data envelopment analysis. *Stata Journal* 10(2): 267–280(14). http://www.stata-journal.com/article.html?article=st0193.

Lovell, C., S. Grosskopf, E. Ley, J. Pastor, D. Prior, and P. Eeckaut. 1994. Linear programming approaches to the measurement and analysis of productive efficiency. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research* 2(2): 175–248. http://ideas.repec.org/a/spr/topjnl/v2y1994i2p175-248.html.

Tauchmann, H. 2012. Partial frontier efficiency analysis. *Stata Journal* 12(3): 461–478(18). http://www.stata-journal.com/article.html?article=st0270.

Thrall, R. M. 1999. What Is the Economic Meaning of FDH? *Journal of Productivity Analysis* 11: 243–50.

About the authors

Byoungin Lim is a software engineer and consultant at NEXTREE SOFT Co., Ltd. in Seoul, Republic of Korea.

Kyoungrok Lee received his Ph.D degree from the Defense Science and Technology Department at Korea National Defense University in Seoul, Republic of Korea.

Choonjoo Lee(Corresponding author) is a professor in the Defense Science and Technology Department at Korea National Defense University in Seoul, Republic of Korea.