Data Envelopment Analysis in Stata

Choonjoo Lee (sarang90@kndu.ac.kr)

Yong-bae Ji (jyb7707@hanmail.net)

Korea National Defense University, Republic of Korea

Stata Conference DC09 (July 30 - 31, 2009)

Outline

1. Why DEA in Stata?
2. The Basics of DEA
3. The Stata/DEA program
4. Stata/DEA Examples
5. DEA Frontiers in Stata?
6. References

1. Why DEA in Stata?

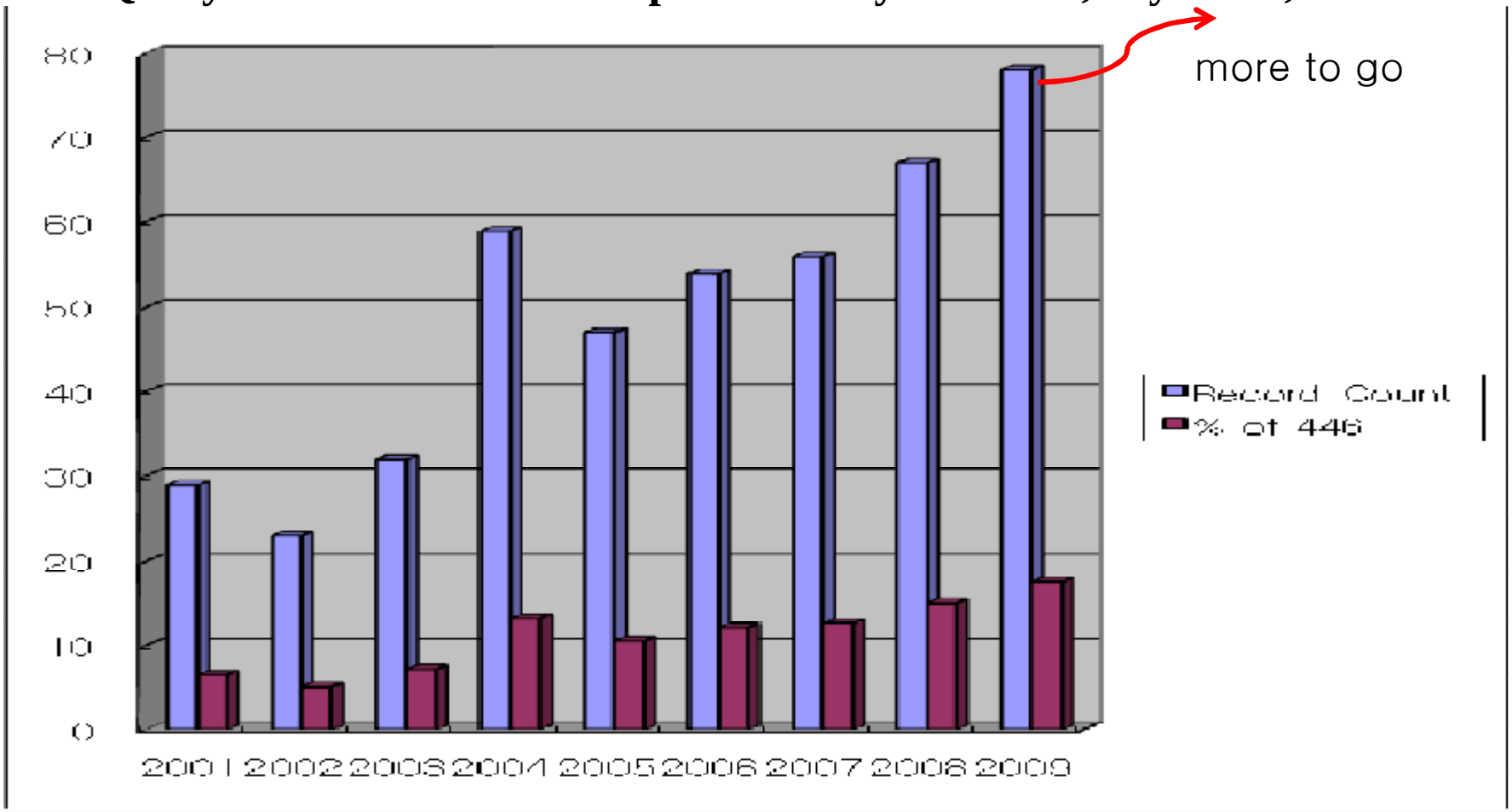
¢ Software Tools for Frontier Analysis

Method	Data Envelopment Analysis	Stochastic Frontier Analysis
Language	AMPL, GAMS, Mathematica, SAS, VBA	SAS, Matlab, R
Program	DEA Excel Solver, DEAP(v 2.1), DEAQual, DEA-Solver-Pro, EMS, FEAR, Frontier Analyst, OnFront, PIM-DEAsoft, Pioneer, Warwick DEA, MaxDEA, KonSi DEA, ISYDS(SIAD), xlDEA, LIMDEP, StoNED	Stata, BSFM, Frontier(v 4.1), WinBUGS, LIMDEP, StoNED
Online Program	DEA Solver Online, iDEA	.

1. Why DEA in Stata? (cont.)

¢ DEA literatures by Year(2001-2009)

☞ **D/B: Science Direct, EBSCO, Google scholar, *SCI, SSCI**
Query: DEA or Data Envelopment Analysis in title, key word, abstract

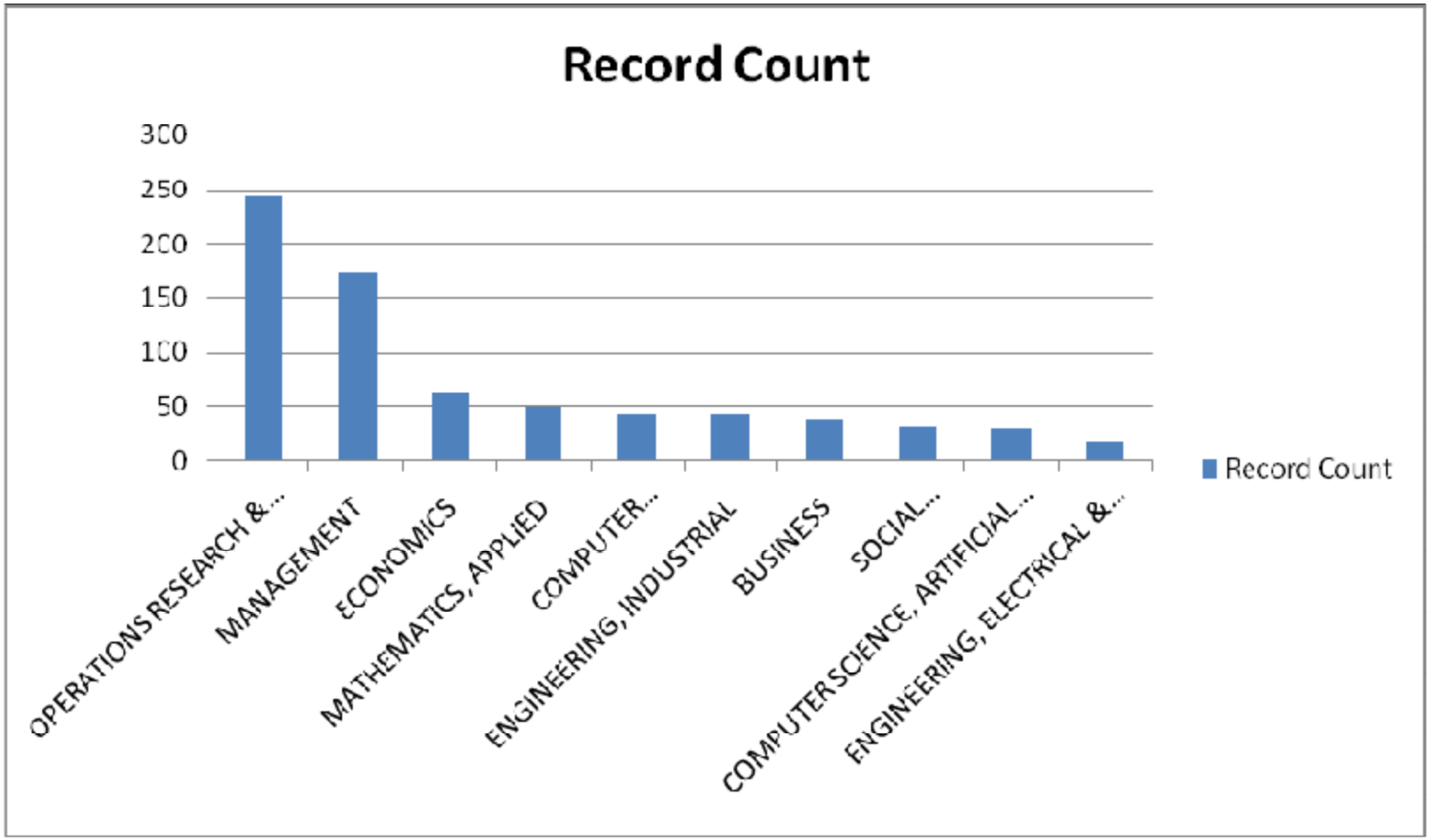


¢ DEA literatures by Journal(2001-2009)

Source Title	Record Count	% of 446
European J. OF Oper. Res.	93	20.8
J. of the Oper. Res. Society	37	8.3
App. Math. and Computation	36	8.1
J. of Prod. Ana.	30	6.7
Omega-Int. of Management Sci.	21	4.7
Comp. & Oper. Research	15	3.3
Expert Systems with Applications	13	2.9
Annal of Oper. Res.	10	2.2
App. Economics	10	2.2
Int. J. of Infor. Tech. & Decision Making	10	2.2

1. Why DEA in Stata? (cont.)

¢ DEA literatures by Subject(2001-2009)

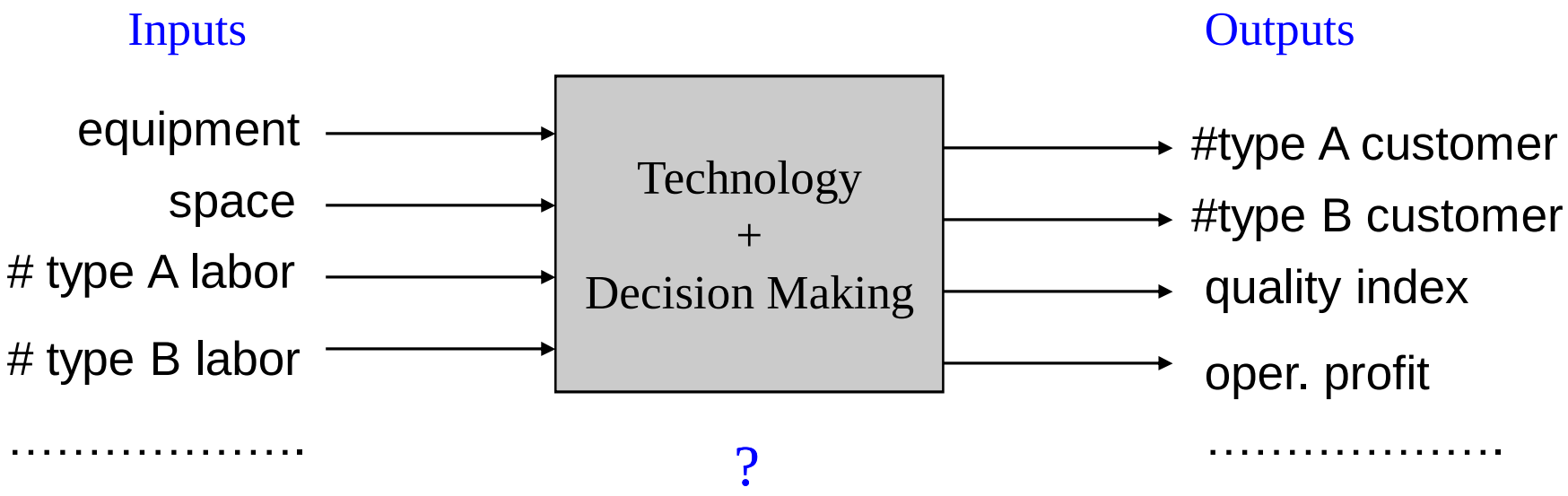


1. Why DEA in Stata? (cont.)

… Stata is easy to use and powerful statistical software; Data Envelopment Analysis code in Stata will promote the efficiency in data management for DEA users and open new application areas in statistical inference for Stata users.

¢ DEA Concept

$$\text{Performance(Efficiency, Productivity)} = \frac{\text{Outputs}}{\text{Inputs}}$$



¢ Assumptions to analyze the black box

- Economic Behaviors: No input, no output!
- (Free) Disposability
- Convexity
- Frontier Search: Piece-wise Linear Method
- Scale Economy
- Orientation: Input-based or Output-based Analysis
- …

¢ Interpretation of DEA Results

- X-inefficiency
- Rational Choice of Input-Output Mixes
- Performance
- …

¢ Terms & Notations

- (X, Y): Input, output matrix
- u, v : Row vector
- $l = (l_1, ..., l_n)^T$: Non-negativity vector
- q, h : Real variable
- Decision Making Units(DMUs)

¢ Basic DEA Models: CCR, BCC

Orientation	Primal	Dual
Input Oriented	$\max \quad z = uy_j - (u_j)^*$ $\text{s.t.} \quad vx_j = 1$ $-vX + uY - (u_j e)^* \le 0$ $v \ge 0, \quad u \ge 0, \quad (u_j \text{ free in sign})^*$	$\min \quad q$ $s.t. \quad q\,x_j - X\,l \ge 0$ $Y\,l \ge y_j$ $(e\,l = 1)^*$ $l \ge 0$
Output Oriented	$\min \quad z = vx_j - (v_j)^*$ $s.t. \quad uy_j = 1$ $vX - uY - (v_j e)^* \ge 0$ $v \ge 0, \quad u \ge 0, \quad (v_j \text{ free in sign})^*$	$\max \quad h$ $s.t. \quad x_j - X\,l \ge 0$ $h\,y_j - Y\,l \le 0$ $(e\,l = 1)^*$ $l \ge 0$

( ) is the additional constraint in BCC model*

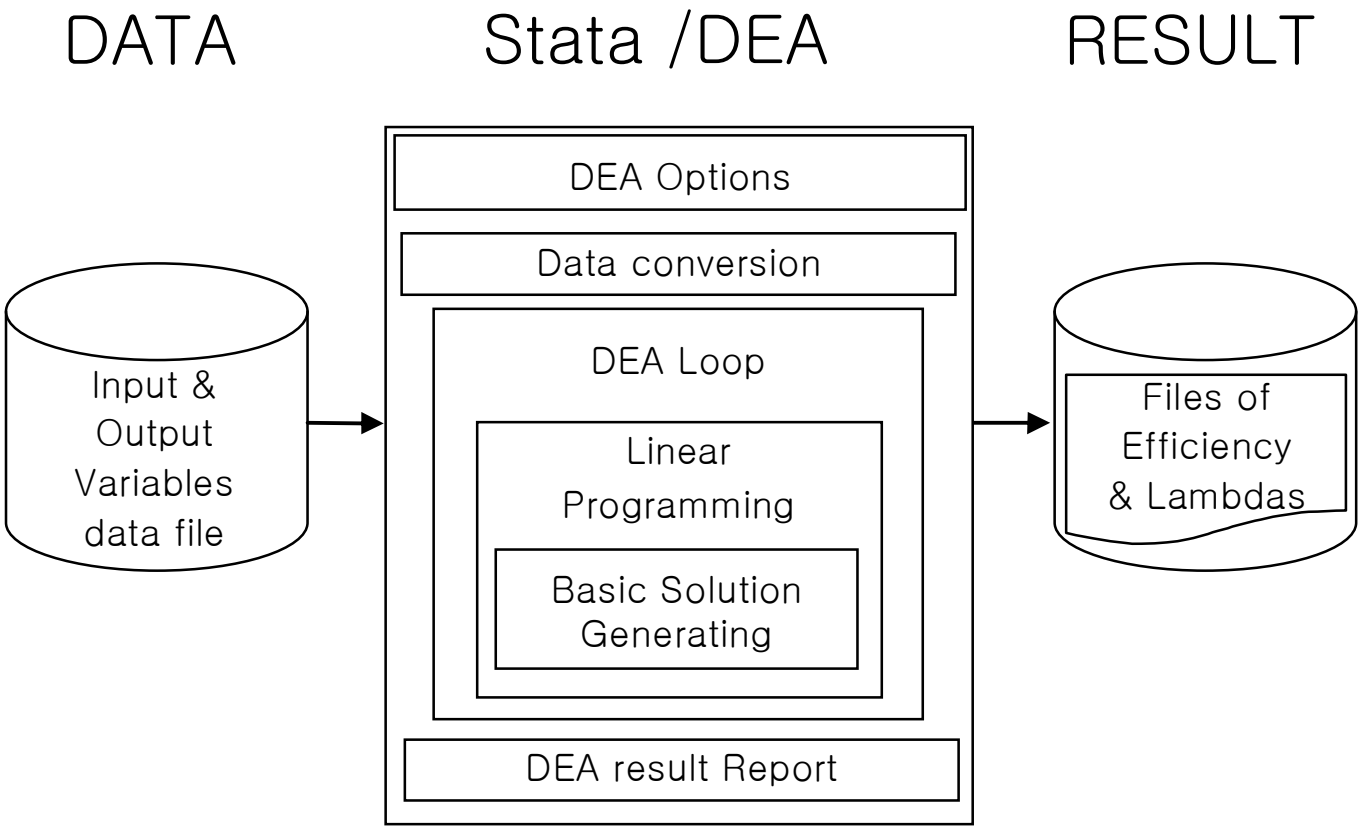
¢ Characteristics of DEA

- No assumption about Input-Output Function
- No limits to the number of inputs and outputs
- Not required to weight restrictions
- Provide reference sets for benchmarking
- Provide useful information for input-output mix decision
- n times computations for n DMUs

¢ User Written Stata/DEA Description

- Considered the basic DEA models (CCR & BCC)
- Can handle both input minimization and output maximization problems
- The data flow in the Stata/DEA program
  - the input and output variables data sets required
  - the DEA options define the model
  - the “Stata/DEA” program consists of “basic” and “lp” subroutine
  - the result data sets available for print or further analysis

¢ Diagram of Data flow in Stata/DEA program



¢ Stata/DEA Syntax (program code under Stata journal review)

- **dea [, data(*string*) iotype(*string*) model(*string*) lambda]**

options	description
data (*filename*)	Specifies the file name of input and output variables for observed units. The data describing inputs/outputs must be presented in the format that variables appear in rows and units in columns and saved as .csv file (Comma delimited MS Office Excel file).
iotype (*input or output*)	Specifies the orientation of DEA model to be solved. " *input* " for input minimization model and " *output* " for output maximization model.
model (*ccr or bcc*)	Specifies the DEA model. " *ccr* " for Charnes-Cooper-Rhodes model and " *bcc* " for Banker-Charnes-Cooper model.
lambda	Lists more information for lambdas and slacks than anyone probably wanted to know.

4. Stata/DEA Examples

¢ **Example 1: Store's efficiency case(for model verification)**

- **Data: two inputs, two outputs, and 5 DMUs**

  ※ *Data imported from Cooper et al.(2006), p.75, Table 3.7*

- **The inputs are**
  - **The number of employees (Employee)**
  - **The floor area (Area)**
- **The outputs are**
  - **The volume of sales (Sales)**
  - **The volume of profits (Profits)**

4. Stata/DEA Examples(cont.)

l The data file including input and output variables

iotype	varname	Store A	Store B	Store C	Store D	Store E
input	Employee	10	15	20	25	12
input	Area	20	15	30	15	9
output	Sales	70	100	80	100	90
output	Profits	6	3	5	2	8

※ The input and output variables are saved in files "ta3_7.csv"

l A user needs to set the options as required and run the following code for input orientation model

- **dea, data(*ta3_7*) iotype(*input*) model(*ccr*) lambda**

4. Stata/DEA Examples(cont.)

¢ The Result Window

```
number of observations will be reset to 5
Press any key to continue, or Break to abort
obs was 0, now 5
c:\data
file eff_ta3_7.dta saved
```

	DMU_No	Score	rank	Refere~1	Refere~2	Refere~3	Refere~4
1.	1	.9333333	2	DMU5			
2.	2	.8888889	3	DMU5			
3.	3	.5333334	5	DMU5			
4.	4	.6666667	4	DMU5			
5.	5	1	1	DMU5			

```
** input_oriented ccr_model DEA results **
```

	optima~s	DMU1	DMU2	DMU3	DMU4	DMU5
1.	.9333333	theta				
2.	.7777778	lambda5				
3.	.2222222	slack4				
4.	11.66667	slack2				
5.	.8888889		theta			
6.	5.888889		slack4			
7.	1.111111		lambda5			
8.	3.333333		slack2			
9.	.5333334			theta		
10.	.8888889			lambda5		
11.	2.111111			slack4		
12.	8			slack2		
13.	.6666667				theta	
14.	1.111111				lambda5	
15.	6.888889				slack4	
16.	3.333333				slack1	
17.	1					theta
18.	1					lambda5
19.	0					slack4
20.	0					slack2

l The result file including the efficiency score and reference set

DMU_No	Score	rank	Reference1	Reference2	Reference3	Reference4
1	0.9333333	2	DMU5			
2	0.8888889	3	DMU5			
3	0.5333334	5	DMU5			
4	0.6666667	4	DMU5			
5	1	1	DMU5			

☞ Scores match with the results of Cooper et. al.(2006).

4. Stata/DEA Examples(cont.)

- **The result file including detail values for the efficiency score and reference set (lambda option)**

optimal_solutions	DMU1	DMU2	DMU3	DMU4	DMU5
0.9333333	theta				
0.7777778	lambda5				
0.2222222	slack4				
11.66667	slack2				
0.8888889		theta			
5.888889					

¢ Example 2: Weapons system construction efficiency

- two inputs, three outputs, and 10 DMUs

※ Data from JAA fr(Jane's Armour and Artillery)

- The inputs are
  - ¡ Combat weight
  - ¡ Height
- The outputs are
  - ¡ Power-to-weight ratio
  - ¡ Max road speed
  - ¡ Main armament diameter

l The data file including input and output variables

iotype	varname	M48A3	M60	T-62	Leopard1A4	T-72	Leopard2	M1	K1	Leclerc	T-90
input	Combat weight	46	40	41	40	41	55.2	54.5	51	54	46.5
input	Height	3.124	3.27	3.33	2.76	2.19	2.64	2.44	2.24	2.46	2.2
output	Power-to-weight ratio	16	16.2	14.5	19.7	19	27	27	23.5	27	17
output	Max road speed	48	48	50	64	80	72	72.4	65	71	60
output	Main armament diameter	90	105	115	105	125	120	105	105	120	125

※ The input and output variables are saved in files "t4_2.csv"

4. Stata/DEA Examples(cont.)

- **dea, data(*t4_2*) iotype(*output*) model(*ccr*) lambda**

```
file lambda_t4_2.dta saved
c:\data\stata_dea
number of observations will be reset to 10
Press any key to continue, or Break to abort
obs was 0, now 10
c:\data
(note: file eff_t4_2.dta not found)
file eff_t4_2.dta saved
```

	DMU_No	Score	rank	Refere~1	Refere~2	Refere~3	Refere~4	Refere~5
1.	1	.716717	10	DMU5	DMU4			
2.	2	.8699198	9	DMU5	DMU4			
3.	3	.92	8	DMU5				
4.	4	1	4	DMU4				
5.	5	1	1	DMU5				
6.	6	.979152	6	DMU4	DMU9			
7.	7	1	3	DMU7				
8.	8	.9735791	7	DMU7	DMU9	DMU5		
9.	9	1	2	DMU9				
10.	10	.9954545	5	DMU5				

```
** output_oriented ccr_model DEA results **
```

	optimal~s	DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7	DMU8	DMU9	DMU10
1.	.716717	theta									
2.	.2775581	lambda5									
3.	.865503	lambda4									
4.	10.6248	slack4									
5.	.1273597	slack2									
6.	.8699198		theta								
7.	.903642		lambda5								
8.	21.83495		slack4								
9.	1.087427		slack2								
10.	.0737669		lambda4								
11.	.92			theta							
12.	1			lambda5							

4. Stata/DEA Examples(cont.)

The result file including the efficiency score and reference set

	DMU_No	Score	rank	Refere~1	Refere~2	Refere~3	Refere~4	Refere~5
1.	1	.716717	10	DMU5	DMU4			
2.	2	.8699198	9	DMU5	DMU4			
3.	3	.92	8	DMU5				
4.	4	1	4	DMU4				
5.	5	1	1	DMU5				
6.	6	.979152	6	DMU4	DMU9			
7.	7	1	3	DMU7				
8.	8	.9735791	7	DMU7	DMU9	DMU5		
9.	9	1	2	DMU9				
10.	10	.9954545	5	DMU5				

** output_oriented ccr_model DEA results **

4. Stata/DEA Examples(cont.)

- **The result file including detail values for the efficiency score and reference set (lambda option)**

	optimal~s	DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7	DMU8	DMU9	DMU10
1.	.716717	theta									
2.	.2775581	lambda5									
3.	.865503	lambda4									
4.	10.6248	slack4									
5.	.1273597	slack2									
6.	.8699198		theta								
7.	.903642		lambda5								
8.	21.83495		slack4								
9.	1.087427		slack2								
10.	.0737669		lambda4								
11.	.92			theta							
12.	1			lambda5							
13.	25.65217			slack4							
14.	3.23913			slack3							
15.	1.14			slack2							
16.	1				theta						
17.	1				lambda4						
18.	0				lambda5						
19.	-4.44e-16				slack2						
20.	0				slack4						
21.	1					theta					
22.	1					lambda5					
23.	-1.42e-14					lambda9					
24.	0					lambda7					
25.	0					slack5					
26.	.979152						theta				
27.	.0837379						lambda4				
28.	.9601942						lambda9				
29.	.0468058						slack2				
30.	1.46076						slack5				
31.	1							theta			
32.	1							lambda7			
33.	0							lambda9			
34.	0							slack5			
35.	0							lambda5			
36.	.9735791								theta		
37.	.2292651								lambda7		
38.	2.313287								slack1		
39.	.5951521								lambda9		
40.	.0988671								lambda5		
41.	1									theta	
42.	1									lambda9	
43.	0									slack2	
44.	0									lambda4	
45.	0									slack4	
46.	.9954545										theta
47.	1.004566										lambda5
48.	20.09132										slack4
49.	2.009132										slack3
50.	5.312786										slack1

5. DEA Frontiers in Stata?

- **The Stata/DEA program is a new application in Stata.**
- **DEA is a prevalent and powerful managerial tool for measuring the performance.**
- **The Stata/DEA program will provide Stata users with several opportunities :**
  - **No extra cost to access DEA**
  - **Flexible DEA model extension and development**
  - **A powerful managerial tool as well as data management, statistical analysis, and optimization procedures**
- **The Stata/DEA program report's files can directly feed to other Stata routines for further analysis.**
- **Further Extensions to 2nd Stage Regression Analysis, DGP of DEA, Statistical Inferences of DEA, Case Specific DEA Models, and more are possible.**

6. References

- **Lee, C., & Ji, Y. (2009). “Data Envelopment Analysis in Stata”, under review by the Stata Journal.**
- **Cooper, W. W., Seiford, L. M., & Tone, A. (2006). Introduction to Data Envelopment Analysis and Its Uses, Springer Science+Business Media.**
- **Charnes, A., Cooper, W. W., & Rhodes, E. (1981). "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through." Management Science, Vol. 27., pp. 668-697.**
- **Banker, R. D., Charnes, A., & Coopers, A. A. (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis”, Management Science Vol. 30, No. 9, pp.1078-1092.**