Data Envelopment Analysis Software MaxDEA X

2026-10-26

MaxDEA Software Ltd.

maxdea@maxdea.com

https://maxdea.com

Contents

ABOUT THIS BOOK 1

CHAPTER 1 BASIC OPERATIONS OF MAXDEA 2

1.1 Main Features of MaxDEA and the Operating Systems Supported 2

1.1.1 Main Features 2

1.1.2 Supported Operating Systems 2

1.2 Basic Operations of MaxDEA 3

1.3 Data Format 4

DMU Names 6

Panel Data Format 6

1.4 Analysis Results 7

CHAPTER 2 BASIC DEA MODELS 8

2.1 The Concept of Technical Efficiency 8

2.2 The CCR Model under Constant Returns to Scale 10

2.2.1 Programming Formulation of the Input-Oriented CCR Model 10

2.2.2 The Nature of the Technical Efficiency Score Obtained from the Input-Oriented CCR Model 15

2.2.3 Programming Formulation of the Output-Oriented CCR Model 16

2.2.4 The Nature of the Technical Efficiency Score Obtained from the Output-Oriented CCR Model 17

2.2.5 Graphical Illustration of the Input-Oriented CCR Model 17

2.2.6 Graphical Illustration of the Output-Oriented CCR Model 19

2.3 The BCC Model under Variable Returns to Scale 21

2.3.1 Programming Formulation of the Input-Oriented BCC Model 21

2.3.2 The Nature of the Technical Efficiency Score Obtained from the Input-Oriented BCC Model 22

2.3.3 Programming Formulation of the Output-Oriented BCC Model 22

2.3.4 The Nature of the Technical Efficiency Score Obtained from the Output-Oriented BCC Model 23

2.3.5 Graphical Illustration of the Input-Oriented BCC Model 23

2.3.6 Graphical Illustration of the Output-Oriented BCC Model 24

2.4 The Comparability of Efficiency Scores among DMUs 25

2.5 The Relationship between Input-Oriented and Output-Oriented Efficiency Scores 26

2.6 The Scale Efficiency Problem 27

2.7 The Naming of DEA Models 29

2.8 Strong Efficiency, Weak Efficiency and the Slack Problem 30

2.9 The Relationship between Radial (Proportional) Improvement and Slack Improvement 35

2.10 Choosing the Number of DMUs, the Input–Output Measures and the Model Orientation 36

2.10.1 The Requirement of DEA Models on the Number of DMUs 36

2.10.2 Choosing the Input and Output Measures 37

2.10.3 Choosing the Model Orientation 39

2.11 A Worked Example: Solving Basic DEA Models in MaxDEA 40

2.11.1 Model Settings *41*
2.11.1.1 Detailed Setup 41
2.11.1.2 Shortcut 42
2.11.2 Analysis Results *42*
2.11.2.1 Efficiency Scores and Slack Variables 43
2.11.2.2 Benchmarks and Projected Values (Target Values) 44
2.11.2.3 Weight Coefficients of the Inputs and Outputs 46
2.11.2.4 Relationship between the Dual Solutions of the Envelopment Model and the Multiplier Model 47
2.11.2.5 Scale Efficiency 47
2.12 Common Misunderstandings about Projected Values 48
2.13 Relationship between the Output Results of the Envelopment Model and the Multiplier Model 49
2.13.1 Graphical Illustration of the Projections of the Envelopment Model and the Multiplier Model *49*
2.13.2 Why the Multiplier Model Has Multiple Optimal Solutions *50*
2.13.3 Why a Weight of Zero Appears in the Multiplier Model Results *50*

CHAPTER 3 RETURNS TO SCALE 51

3.1 Types of Returns to Scale in DEA Models 51
3.2 Determining the RTS Status of a DMU 54
3.2.1 Determining RTS Status through the Envelopment Model *54*
3.2.2 Determining RTS Status through the Multiplier Model *56*
3.3 Graphical Illustration of Solving for Scale Efficiency and RTS Status in MaxDEA 61
3.4 The Relationship between RTS Status and Model Orientation 61

CHAPTER 4 DISTANCE FUNCTIONS IN DEA 62

4.1 Radial Distance 62
4.2 Farthest Distance to the Frontier (SBM Model) 64
4.2.1 The SBM Model *64*
4.2.2 The Weighted SBM(ERM) Model *67*
4.2.3 The MSBM Model *68*
4.2.4 A Worked Example: Solving the SBM(ERM) Model and the Weighted SBM(ERM) Model in MaxDEA *69*
4.3 Minimum Distance to the Weak Efficient Frontier 72
4.3.1 Computing the Minimum Distance to the Weak Efficient Frontier *72*
4.3.2 A Worked Example: Solving the MinDW Model in MaxDEA *73*
4.4 Minimum Distance to the Strong Efficient Frontier 75
4.4.1 Computing the Non-Oriented Minimum Distance to the Strong Efficient Frontier *75*
4.4.2 A Worked Example: Solving the Non-Oriented MinDS Model in MaxDEA *88*
4.4.2.1 The CRS Model of Minimum Distance to the Strong Efficient Frontier 89
4.4.2.2 The VRS Model of Minimum Distance to the Strong Efficient Frontier 90
4.4.3 Peculiarities of the MinDS Model *92*
4.4.3.1 The MinDS CRS Efficiency Score May Exceed the VRS Efficiency Score 92
4.4.3.2 Expansion of the Reference Set May Increase the MinDS Efficiency Score 93
4.4.3.3 The Monotonicity Problem 96
4.4.3.4 Input- and Output-Oriented MinDS Models 96

4.4.4 Recent Advances in Algorithms for the MinDS Model *99*
4.5 Directional Distance Function 101
4.5.1 The Directional Distance Function Model *101*
4.5.2 The Multiplier Form of the Directional Distance Function Model *103*
4.5.3 Treatment of Undesirable Outputs in the Directional Distance Function Model and Whether Weak Disposability Is Reasonable *104*
4.5.4 Some Special Cases of the Directional Distance Function Model *113*
4.5.5 Computing the Efficiency Score of a General Directional Distance Function Model *116*
4.5.6 A Worked Example: Solving the Directional Distance Function Model in MaxDEA *118*
4.5.6.1 An Example of the Directional Distance Function Model 118
4.5.6.2 An Example of the Directional Distance Function Model with Undesirable Outputs 122
4.5.7 A Common Misunderstanding about the Results of DEA Models with Undesirable Outputs *125*
4.6 Weighted Additive Distance (Including SBM Directional Distance Function and Non-radial Directional Distance Function) 130
4.6.1 Unweighted Additive Distance (Simple Additive Distance) *130*
4.6.2 Weighted Additive Distance *131*
4.6.3 A Simple-Weight Variant: Slacks-based Measure on the Sum of Directional Distance Functions *134*
4.6.4 The RAM Model of Weighted Additive Distance *135*
4.6.5 The BAM Model of Weighted Additive Distance *137*
4.6.6 The Directional SBM of Weighted Additive Distance (Directional Slacks-based Measure) *138*
4.6.7 The Non-radial Directional Distance Function of Weighted Additive Distance *142*
4.6.8 Relationships among the Various Weighted Additive Distances *148*
4.7 The EBM Hybrid Distance 150
4.7.1 The EBM Model *150*
4.7.2 Shortcomings of the EBM Model and Methods of Improvement *153*
4.7.3 Solving Hybrid Distance Function Models in MaxDEA *161*
4.8 Distances That Incorporate Price Information 161
4.8.1 Cost Efficiency and Revenue Efficiency Models *161*
4.8.2 Solving Cost Efficiency and Related Models in MaxDEA *163*
4.9 Analysing the Degree of Inefficiency of Individual Input and Output Measures 164

CHAPTER 5 SUPER-EFFICIENCY MODELS 166

5.1 Radial Super-efficiency Model 167
5.2 Directional Distance Function Super-efficiency Model 170
5.3 SBM(ERM) Super-efficiency Model 171
5.4 Weighted Additive Distance Super-efficiency Model 177
5.5 Projection Analysis in Super-efficiency Models 178
5.6 Detecting Outliers with Super-efficiency Models 179

CHAPTER 6 OTHER MODELS AND METHODS 181

6.1 The Multiplier Ratio Constraint Model 181
6.1.1 Constraining the Weight Ratio of Two Inputs *181*

6.1.2 Constraining the Weight Ratio of an Output and an Input ... *182*
6.1.3 Constraining the Weight Ratio of One Output and Another Output ... *183*
6.2 The FDH Model ... 184
6.3 Models Containing Undesirable Outputs ... 186
6.3.1 Radial Model Containing Undesirable Outputs ... *186*
6.3.2 SBM Model Containing Undesirable Outputs ... *186*
6.3.3 Directional Distance Function Model Containing Undesirable Outputs ... *188*
6.3.4 Weighted Additive Model Containing Undesirable Outputs ... *188*
6.3.4.1 Simple Additive Model Containing Undesirable Outputs ... 188
6.3.4.2 Weighted Additive Model Containing Undesirable Outputs ... 188
6.3.5 Handling Undesirable Outputs by Means of Negative Measures ... *188*
6.4 Handling Negative Values in MaxDEA ... 192
6.5 External Uncontrollable Factors or Uncontrollable Input/Output Measures ... 192

CHAPTER 7 GENERALIZED DEA MODELS ... 194

7.1 The Generalized DEA Model ... 194
7.2 Group Benchmarking Models ... 197
7.2.1 Self-benchmarking ... *197*
7.2.2 Cross-benchmarking ... *198*
7.2.3 Downward-benchmarking ... *198*
7.2.4 Upward-benchmarking ... *199*
7.2.5 Lower-adjacent Benchmarking ... *200*
7.2.6 Upper-adjacent Benchmarking ... *200*
7.2.7 Window Benchmarking ... *201*
7.2.8 Fixed Benchmarking ... *203*
7.3 Data Format for Group Benchmarking Models ... 204

CHAPTER 8 PANEL DATA MODELS ... 206

8.1 Panel Data ... 206
8.2 The Relationship between Productivity Change, Technological Change and Technical Efficiency Change ... 211
8.3 Basic Principles of the Malmquist Index ... 214
8.4 Methods for Computing the Various Types of Malmquist Index ... 218
8.4.1 MI for Different Periods Computed against the Frontiers of Different Periods ... *218*
8.4.1.1 Adjacent Reference (Average of Two Indices) ... 218
8.4.1.2 Sequential Reference (Average of Two Indices) ... 221
8.4.1.3 Window Reference (Average of Two Indices) ... 224
8.4.1.4 Further Decomposition of TC ... 225
8.4.1.5 The Infeasibility Problem of the VRS Model Caused by Cross-Period Benchmarking ... 227
8.4.1.6 Biennial Frontier (Single Index) ... 227
8.4.1.7 Further Decomposition of TC under the Biennial Frontier ... 229
8.4.1.8 The Non-multiplicability Problem of the MI ... 230
8.4.2 Computing the MI of All Periods against the Same Frontier ... *230*

8.4.2.1 Fixed Benchmarking (Single Index) 230
8.4.2.2 Global Benchmarking (Single Index) 231
8.4.2.3 Fixed Window Benchmarking 233
8.4.2.4 Further Decomposition of TC under the Same Frontier 233
8.5 Comparison of the Various MIs 235
8.6 Malmquist Index Decomposition Methods Commonly Used in the Literature 240
8.6.1 R Färe et al. (1992) Decomposition Method 240
8.6.2 R. Färe et al. (1994) Decomposition Method 240
8.6.3 Ray & Desli (1997) Decomposition Method 241
8.6.4 Zofio (2007) Decomposition Method 241
8.7 The Luenberger Index 242
8.8 The Malmquist-Luenberger Index 242
8.9 The Cost-Malmquist Index 243
8.10 The Hicks-Moorsteen TFP Index (HMB Index) 243
8.11 Some Suggestions for Panel Data Analysis 244
8.12 The Window DEA Model for Panel Data 244
8.13 How to Analyse the Determinants of TFP When TFP Cannot Be Measured 244

CHAPTER 9 INNOVATIVE TWO-STAGE AND GENERALIZED NETWORK DEA 247

Notation and Data Description 248
9.1 Two-Stage Network DEA (Constant Returns to Scale, CRS) 249
9.1.1 The Two-Stage Network DEA Model *249*
9.1.2 Two-Stage Network DEA Model Based on Conservation of the Virtual Value of Intermediates *250*
9.1.3 Efficiency Decomposition of the Two-Stage Network DEA Model (CRS) *252*
9.1.4 The Flexible Synergy Mechanism for Intermediates *252*
9.1.5 Why the Efficiency Decomposition of Two-Stage Network Models Takes the Product Form *254*
9.2 Two-Stage Network DEA (Variable Returns to Scale, VRS) 255
9.2.1 Model Definitions *255*
9.2.2 Efficiency Decomposition *256*
9.3 Generalized Network DEA Model 257
9.3.1 Model Definition *257*
9.3.2 Solving the Model *257*
9.3.3 Method for Computing System Efficiency *258*
9.3.4 Compatibility with the Preceding Two-Stage Model *259*
9.4 Generalized Network DEA Based on Full-Dimensional Efficient Facets (FDEF) 261
9.4.1 Introduction to Full-Dimensional Efficient Facets (FDEF) *262*
9.4.2 Generalized Network DEA Model Based on the Affine Hull of FDEFs *263*
9.5 Software Demonstration of Network DEA 264
9.5.1 Defining the Data *264*
9.5.2 Setting the General Options of the DEA Model *264*
9.5.3 Setting Other Model Options *264*

9.5.4 Defining the Network Structure *265*
9.5.5 Results *265*
APPENDIX: PROOF OF THE MULTIPLICATIVE DECOMPOSITION OF EFFICIENCY IN THE CRS TWO-STAGE NETWORK DEA 266
REFERENCES 268

About This Book

Data envelopment analysis (DEA) is a very widely applied method of efficiency analysis. By combining methodological exposition with software operation, and mathematical models with graphical illustrations, this book aims to provide the many users of DEA with a guide that is easy to understand, simple and practical.

Chapter 1 introduces the basic operations of MaxDEA, Chapter 2 the basic DEA models (CCR and BCC), Chapter 3 the types of returns to scale in DEA models and methods for calculating scale elasticity, Chapter 4 the various distance functions used in DEA models (including undesirable outputs), and Chapter 5 the super-efficiency models (including undesirable outputs), together with a solution to the infeasibility problem of the VRS super-efficiency model. Chapter 6 briefly introduces other DEA models, Chapter 7 deals with generalized DEA models and Chapter 8 with panel data models. Chapter 9 presents the new two-stage network model and the generalized network model developed by the MaxDEA team. Each chapter contains either a worked example solved in MaxDEA or instructions on how to operate the software.

Chapter 1 Basic Operations of MaxDEA

1.1 Main Features of MaxDEA and the Operating Systems Supported

1.1.1 Main Features

1) Simple to use, with native cross-platform support for Windows, macOS and Linux.
2) Supports DEA models with large data sets, and solves them extremely fast.
3) Full support for the latest advances in the MinDS model: the strongly monotonic algorithm; CRS, VRS, NIRS and NDRS; and non-oriented, input-oriented and output-oriented models.
4) The most important feature of MaxDEA is that it contains a large number of DEA methods and, as far as possible, offers combined applications among the various DEA methods. A combined model is built simply by setting several model options at the same time. For example, the cost model combined with the Malmquist model gives the cost Malmquist model; the undesirable output model combined with the directional distance function model and the Malmquist model gives a directional distance function Malmquist model that incorporates undesirable outputs (this combination is usually called the Malmquist-Luenberger index); and so on.

1.1.2 Supported Operating Systems

1) Windows 7 (64-bit) or later.
2) macOS 10.11 or later.
3) The various Linux distributions (x86-64).

1.2 Basic Operations of MaxDEA

MaxDEA provides a simple and clear navigation bar: Step 1 - Step 2 - Step 3.



Figure 1-1 MaxDEA Navigation Bar

Step 1 is to create a new blank MaxDEA model file.

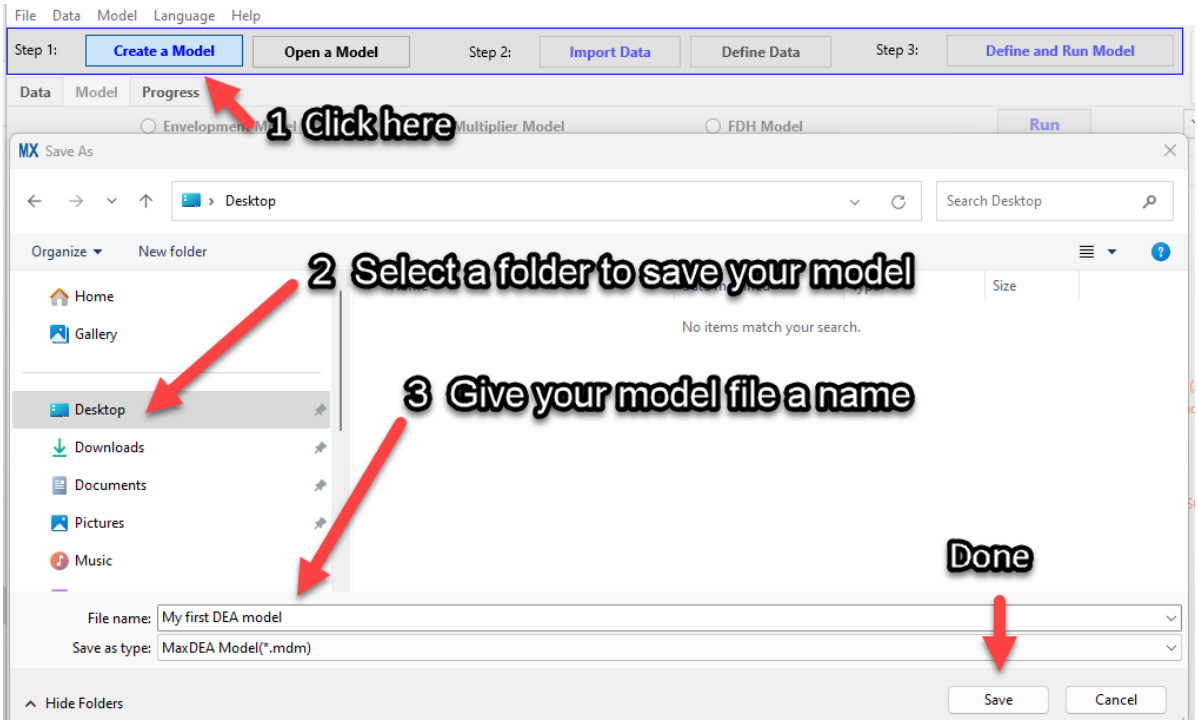


Figure 1-2 Creating a New Model

Step 2 is to import and define the data.

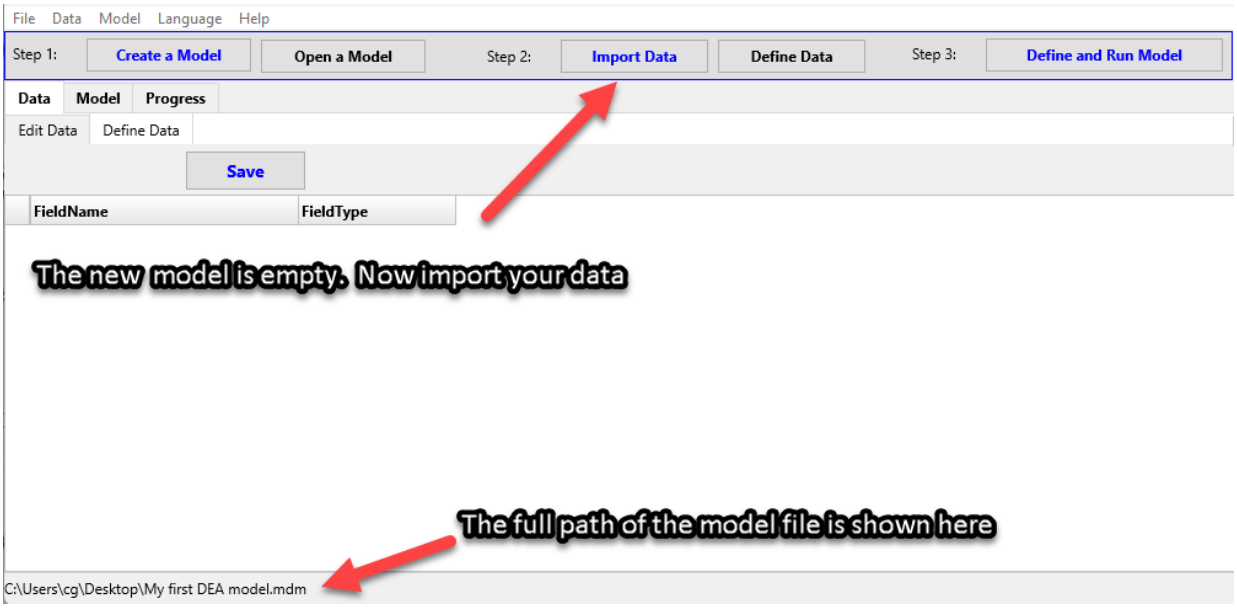


Figure 1-3 Importing Data

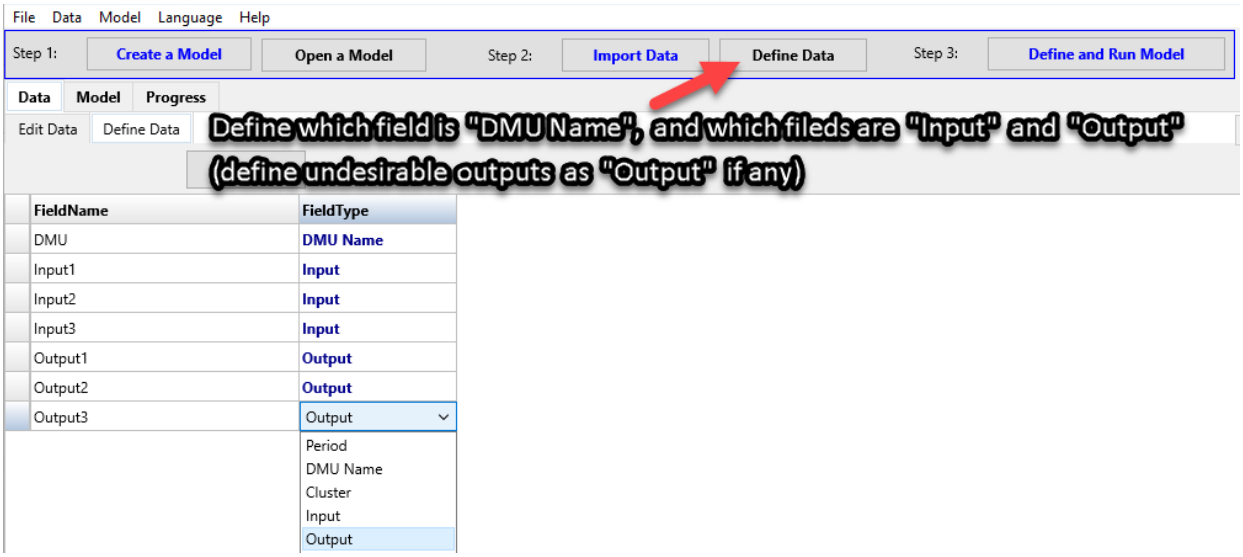


Figure 1-4 Defining Data

Step 3 is to set up and run the model.

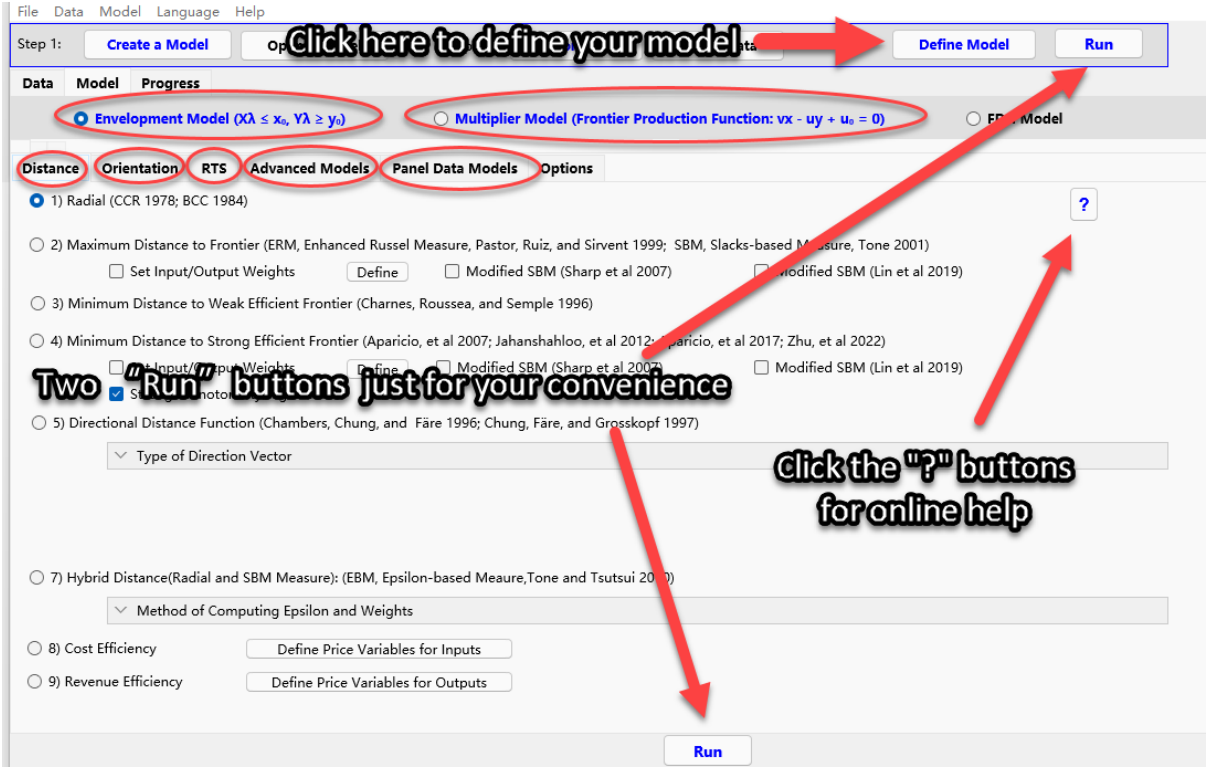


Figure 1-5 Setting Up and Running the Model

1.3 Data Format

The data types supported by MaxDEA include:

1) Excel files
2) Open Office Calc files (*.ods; *.ots)
3) Text files (*.csv; *.tsv; *.ssv; *.psv; *.txt; *.zip)

   For text files, the data format must be indicated by the file extension:

   *.csv: comma-separated text file (the most commonly used text data format)

   *.tsv: tab-separated text file

   *.ssv: semicolon (;) separated text file

   *.psv: vertical bar (|) separated text file

   For text files with the .txt or .zip extension, the user can specify the format when importing the data. A file with the .zip extension is a compressed .txt or .csv file.
4) Stata (*.dta)
5) SPSS (*.sav; *.zsav)
6) SAS (*.sas7bdat; *.xpt; *.v8xpt)
7) dBase, FoxPro (*.dbf)
8) Access databases
9) SQLite databases (*.sdb)

If the data file is a spreadsheet (Excel, Open Office Calc or a text file), the first row of the data must contain the variable names (Chinese is supported), and only one row may give the variable names. MaxDEA imposes no special requirements on the names of the variables or on the order in which they are arranged. The variable names of the input (output) variables can be chosen freely.

Table 1-1 Correct Data Format

Company	Capital	Labor	Product
A	4323	875	93608
B	2295	469	225559
C	6379	1286	327068
D	6644	1339	201354
E	1436	297	188926

Table 1-2 Incorrect Data Format

DMU	Input		Output
Company	Capital	Labor	Product
A	4323	875	93608
B	2295	469	225559
C	6379	1286	327068
D	6644	1339	201354
E	1436	297	188926

When the volume of data is large, erroneous or anomalous data are often present. Before importing the data, the data should be cleaned with appropriate data-processing software (for example, statistical analysis software). Excel is merely a spreadsheet and has no strict data types, so errors that are difficult to detect can easily arise; Excel is therefore not suitable for data cleaning. MaxDEA X supports input and output variable values of 0. If there are blank values among the input and output variables, MaxDEA X automatically treats them as 0.

DMU Names

The data table must contain a column that gives the name of each DMU (DMU Name). A DMU name may be textual (Chinese is supported), numeric, or a combination of letters and numbers.

Panel Data Format

In panel data, one column must indicate the time period of the data. MaxDEA imposes no special requirements on the ordering of panel data. Panel data may be sorted by DMU name, may be sorted by time period, or may not be sorted at all.

The variable that indicates time must be an integer; it may be a consecutive integer sequence, for example

1, 2, 3……

2001, 2002, 2003……

or a non-consecutive integer sequence, for example

1, 2, 5, 8……

2001, 2005, 2009……

200101, 200102, 200302……

MaxDEA allows unbalanced panel data, that is, it permits the data for some periods to be missing (missing as an entire record). For the periods for which data are missing, the corresponding analysis results are also missing.

Table 1-3 Example of a Panel Data Format (Balanced Panel)

Period	Company	Capital	Labor	Product
1	A	4323	875	93608
1	B	2295	469	225559
1	C	6379	1286	327068
2	A	6644	1339	201354
2	B	1436	297	188926
2	C	6281	1266	413738
3	A	7459	1502	114022
3	B	4464	903	212444
3	C	4524	915	462677

Table 1-4 Example of a Panel Data Format (Unbalanced Panel)

Period	Company	Capital	Labor	Product
1	A	4323	875	93608
1	B	2295	469	225559
1	C	6379	1286	327068
2	A	6644	1339	201354
2	B	1436	297	188926
3	A	7459	1502	114022
3	B	4464	903	212444
3	C	4524	915	462677

1.4 Analysis Results

The analysis results are saved in a separate folder in the same location as the model file, and the folder is named after the current date. If that folder already exists, "current date plus a numeric suffix" is used. When the software has finished running, it displays the location where the results are saved.

Chapter 2 Basic DEA Models

2.1 The Concept of Technical Efficiency

Technical efficiency refers to the extent to which the production process of a production unit reaches the technological level of the industry. Technical efficiency reflects how high or low the technological level of a production unit is, and it is for this reason that it is called "technical" efficiency. Technical efficiency can be measured from two perspectives, input and output: given the inputs, technical efficiency is measured by the extent to which output is maximized; given the outputs, technical efficiency is measured by the extent to which input is minimized(Farrell, 1957).

When the influence of the scale of production is not considered, technical efficiency can be measured quantitatively by the ratio of output to input, that is, expressed as y/x. If the influence of the scale of production is disregarded and multiple periods are not involved, technical efficiency and productivity express the same meaning. When the production process involves only one input and one output, the output/input ratio of each production unit, that is, the quantity of output produced per unit of input consumed, can be calculated in order to reflect the level of technical efficiency of each production unit. If the output/input ratio of each unit is divided by the maximum of these ratios, the output/input ratios can be standardized to values between 0 and 1, which reflects more clearly the gap in technical efficiency between the evaluated unit and the best unit (Table 2-1).

Table 2-1 Measuring Technical Efficiency with the Output/Input Ratio

Unit	x (Input)	y (Output)	y/x	y/x (Standardized)
A	10	8	0.800	1.000
B	15	12	0.800	1.000
C	32	16	0.500	0.625
D	48	18	0.375	0.469
E	24	12	0.500	0.625
F	54	20	0.370	0.463
G	50	20	0.400	0.500

Although the above method is simple, it is applicable only to the case of a single input and a single output. If the production process involves more than one input or output, a single ratio cannot be calculated directly. In the case of multiple inputs and multiple outputs, there are two approaches to calculating the output/input ratio.

The first is to assign certain weights to the input and output measures and then to calculate the ratio of weighted output to weighted input, which can serve as an index reflecting technical efficiency. This is the approach of the DEA multiplier model.

The second is to take a set of benchmark input-output data (x*, y*) as the reference, to compare the

input-output data of the evaluated unit (xk, yk) with it, and to take the ratio to the benchmark as the technical efficiency of the evaluated data, expressed as $E_{(x_k,y_k)}: E_{(x^*,y^*)} = \frac{y_k / y^*}{x_k / x^*}$ This is the approach of the DEA envelopment model.

The linear programming models constructed by these two approaches are mutually dual, and the technical efficiency scores they yield are equal.

The first approach is taken as the starting point for understanding the basic principle of DEA.

Suppose there are m inputs and q outputs; the weighted input is then expressed as

$v = v_1x_1 + v_2x_2 + \ldots + v_mx_m$ (2.1)

the weighted output is expressed as

$u = u_1y_1 + u_2y_2 + \ldots + u_qy_q$ (2.2)

The next question is how to determine the weight coefficients that reflect the relative importance of the various inputs and outputs. One method is to adopt fixed weights, for example by determining the weight of each measure subjectively through expert consultation or discussion; the other method is to obtain the input and output weights from the data themselves, and the DEA multiplier model adopts this latter method.

2.2 The CCR Model under Constant Returns to Scale

In 1978, Charnes, Cooper and Rhodes published the paper "Measuring the efficiency of decision making units" in the European Journal of Operational Research, thereby founding the DEA theoretical method(A Charnes, Cooper, & Rhodes, 1978). In the subsequent DEA literature, the first DEA model they created is named after the initial letters of the surnames of Charnes, Cooper and Rhodes, namely the CCR model. The CCR model assumes constant returns to scale (CRS); the technical efficiency it yields contains a component of scale efficiency and is therefore usually called overall technical efficiency.

DEA calls the objects whose efficiency is measured decision making units (DMUs). A DMU may be any department or unit with measurable inputs and outputs, such as a firm, a school, a hospital or a project implementation unit (region); it may also be an individual, such as a teacher, a student or a doctor. DMUs must be comparable with one another.

2.2.1 Programming Formulation of the Input-Oriented CCR Model

The previous section discussed the measurement of technical efficiency by calculating the output/input ratio; this topic is now continued in order to expound the basic principle of DEA. Suppose that the technical efficiency of a group of n DMUs is to be measured, denoted DMUj (j=1,2,...,n); each DMU has m inputs, denoted xi (i=1,2,...,m), the input weights being denoted vi (i=1,2,...,m); and q outputs, denoted yr (r=1,2,...,q), the output weights being denoted ur (r=1,2,...,q). The DMU currently being measured is denoted DMUk, and its output/input ratio is expressed as

$$h_k = \frac{u_1 y_{1k} + u_2 y_{2k} + \ldots + u_q y_{qk}}{v_1 x_{1k} + v_2 x_{2k} + \ldots + v_m x_{mk}} = \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}$$

$$v \geq 0; u \geq 0$$

$$i = 1, 2 \ldots, m; r = 1, 2 \ldots, \mathrm{q} \quad (2.3)$$

A further condition is now attached to the technical efficiency score to be measured: the efficiency scores θj obtained by all DMUs under the above weights are restricted to the interval [0, 1], that is,

$$h_k = \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}$$

$$\frac{\sum_{r=1}^{q} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1$$

$$v \ge 0; u \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,\mathrm{q}; j = 1,2\ldots,n \quad (2.4)$$

The first DEA model created by Charnes, Cooper and Rhodes is based on constant returns to scale and is expressed as

$$\max \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}}$$

$$s.t. \ \frac{\sum_{r=1}^{q} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1$$

$$v \ge 0; u \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.5)$$

The meaning of this nonlinear programming model is that, subject to the condition that the efficiency score of no DMU exceeds 1, the efficiency score of the evaluated DMU is maximized. The weights u and v determined by the model are therefore the most favourable to the evaluated DMUk. In this sense, the CCR model gives a conservative estimate of the inefficiency of the evaluated DMU, because the weights it adopts are those most favourable to the evaluated unit, and the efficiency score obtained under any other set of weights will not exceed the score obtained under this set.

Why is the CCR model said to be based on constant returns to scale? If a production technology exhibits constant returns to scale, then, with technical efficiency held unchanged, an increase of the inputs of a DMU to t times their original level (t>0) will be accompanied by a corresponding increase of its outputs to t times their original level. Conversely, suppose that both the inputs and the outputs of the evaluated DMUk become t times their original values; under the constant-returns-to-scale assumption, its technical efficiency should remain unchanged. Is the technical efficiency obtained by the CCR model in this case in fact unchanged?

After the inputs and the outputs of DMUk have both become t times their original values, the objective function of the CCR model is

$$\max \frac{\sum_{r=1}^{q} u_r t y_{rk}}{\sum_{i=1}^{m} v_i t x_{ik}} = \frac{t\sum_{r=1}^{q} u_r y_{rk}}{t\sum_{i=1}^{m} v_i x_{ik}} = \frac{\sum_{r=1}^{q} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \quad (2.6)$$

Since t appears in both the numerator and the denominator, it cancels out and the expression is equivalent to the original objective function. By the same reasoning, the constraints corresponding to DMUk are equivalent to the original constraints. Hence, when both the inputs and the outputs of the evaluated DMUk increase to t times their original values, the CCR model is equivalent to the original model and the efficiency score obtained is unchanged, which is consistent with the constant-returns-to-scale assumption.

The CCR model(2.5) has the problem that it is a nonlinear programming problem and that infinitely many optimal solutions exist. Suppose that the vectors u* and v* constitute an optimal solution of the model(2.5)[i], then tu* and tv* are certainly also optimal solutions of the model (t>0).

Since $\sum_{i=1}^{m} v_k x_{ij} > 0$, the constraints of the model(2.5) are equivalent to

$$s.t. \sum_{r=1}^{q} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \quad (2.7)$$

Let $t = \frac{1}{\sum_{i=1}^{m} v_i x_{ik}}$, then

The objective function of the model(2.5) becomes

$$\max t\sum_{r=1}^{q} u_r y_{rk} = \sum_{r=1}^{q} t u_r y_{rk} \quad (2.8)$$

Furthermore, let $\mu = tu$, $\nu = tv$[ii]

The nonlinear model(2.5) is transformed into an equivalent linear programming model

[i]In this book the solution of a variable in a linear program is denoted by adding an asterisk superscript to that variable.

[ii]Note that the Greek letters μ and ν are used before the equals sign and the English letters u and v after it. Different letters are used because the transformed model and the preceding model are two different linear programs, even though the optimal values of their objective functions are equal.

$$\max \sum_{r=1}^{q} \mu_r y_{rk}$$

$$s.t. \sum_{r=1}^{q} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0$$

$$\sum_{i=1}^{m} v_i x_{ik} = 1$$

$$v \ge 0; \mu \ge 0$$

$$i = 1, 2 \ldots, m; r = 1, 2 \ldots, q; j = 1, 2 \ldots, n \quad (2.9)$$

The model(2.9) is the linear programming formulation of the input-oriented CCR model, written for the case of solving for DMU k; a separate programming formulation must be established for each DMU.

Table 2-2 Example Data for the Linear Programming Formulation

DMU	x1	x2	y
A	10	40	10
B	15	25	10
C	54	27	18
D	50	60	20

Using the data in Table 2-2 as an example, 4 linear programming models must be solved separately; for A through D they are, in order,

DMU A

max +10 Weight_y

s.t. +10 Weight_x1 +40 Weight_x2 = 1

-10 Weight_x1 -40 Weight_x2 +10 Weight_y <= 0

-15 Weight_x1 -25 Weight_x2 +10 Weight_y <= 0

-54 Weight_x1 -27 Weight_x2 +18 Weight_y <= 0

-50 Weight_x1 -60 Weight_x2 +20 Weight_y <= 0

DMU B

max +10 Weight_y

s.t. +15 Weight_x1 +25 Weight_x2 = 1

-10 Weight_x1 -40 Weight_x2 +10 Weight_y <= 0

-15 Weight_x1 -25 Weight_x2 +10 Weight_y <= 0

-54 Weight_x1 -27 Weight_x2 +18 Weight_y <= 0

-50 Weight_x1 -60 Weight_x2 +20 Weight_y <= 0

DMU C

max +18 Weight_y

s.t. +54 Weight_x1 +27 Weight_x2 = 1

-10 Weight_x1 -40 Weight_x2 +10 Weight_y <= 0

-15 Weight_x1 -25 Weight_x2 +10 Weight_y <= 0

-54 Weight_x1 -27 Weight_x2 +18 Weight_y <= 0

-50 Weight_x1 -60 Weight_x2 +20 Weight_y <= 0

DMU D

max +20 Weight_y

s.t. +50 Weight_x1 +60 Weight_x2 = 1

-10 Weight_x1 -40 Weight_x2 +10 Weight_y <= 0

-15 Weight_x1 -25 Weight_x2 +10 Weight_y <= 0

-54 Weight_x1 -27 Weight_x2 +18 Weight_y <= 0

-50 Weight_x1 -60 Weight_x2 +20 Weight_y <= 0

The dual of the model(2.9) is[i]

$$\min \theta$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rk}$$

$$\lambda \ge 0$$

$$i = 1, 2 \ldots, m; r = 1, 2 \ldots, q; j = 1, 2 \ldots, n \quad (2.10)$$

In the dual model(2.10), λ denotes the linear combination coefficient of the DMUs, the optimal solution θ* of the model represents the efficiency score, and θ* lies in the range (0, 1].

($x = \sum_{j=1}^{n} \lambda_j x_j$, $y = \sum_{j=1}^{n} \lambda_j y_j$) can be regarded as a virtual DMU whose inputs are no higher than those of DMUk and whose outputs are no lower than those of DMUk. If DMUk is inefficient, then the virtual DMU constructed from the optimal solution ($\hat{x} = \sum_{j=1}^{n} \lambda_j^* x_j$, $\hat{y} = \sum_{j=1}^{n} \lambda_j^* y_j$) is the projected value of the evaluated DMUk.

The optimal solution of the objective function of the model is θ*. 1- θ* represents the maximum extent to which the inputs of the evaluated DMUk can be reduced at the current technology level without lowering its output level. The smaller θ* is, the greater the extent to which inputs can be reduced and the

[i]For the relationship between a primal model and its dual, see the relevant textbooks, for example Hillier, F. S., & Lieberman, G. J. (2010). Introduction to Operations Research: McGraw-Hill Higher Education.

lower the efficiency. $\theta^* = 1$ indicates that the evaluated DMU lies on the frontier, that none of its inputs can be reduced proportionally without reducing output, and hence that it is technically efficient; $\theta^* < 1$ indicates that the evaluated DMU is technically inefficient, and that each of its inputs can be reduced proportionally by (1- θ^*) without reducing output.

In the model(2.9), the input and output weight coefficients and the inputs and outputs themselves stand formally in the relation of multipliers and multiplicands, and the model(2.9) is usually called the multiplier form of DEA. The frontier determined by its dual model(2.10) is shaped like an envelope that wraps all DMUs inside (see the graphical illustration below), and is usually called the envelopment form of DEA.

The CCR dual model(2.10) measures inefficiency by the extent to which the various inputs can be reduced proportionally while the outputs are held fixed, and is therefore called the input-oriented CCR model.

2.2.2 The Nature of the Technical Efficiency Score Obtained from the Input-Oriented CCR Model

It was mentioned above that two approaches exist for measuring technical efficiency in the case of multiple inputs and multiple outputs, the second of which is the benchmarking method.

The model(2.10) can be equivalently expressed as

$$\min \theta$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rk}$$

$$\lambda, s^-, s^+ \geq 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.11)$$

In this model, the benchmark can be understood as $\left(x^*, y^*\right) = (\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^-, \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+)$, in which case the ratio of the output/input ratio of the evaluated DMU to the output/input ratio of the benchmark (that is, the ratio of the two ratios) is

$$\frac{y_k / x_k}{y^* / x^*} = \frac{y_k / y^*}{x_k / x^*} = \frac{y_k / (\sum_{j=1}^{n} \lambda_j y_j - s^+)}{x_k / (\sum_{j=1}^{n} \lambda_j x_j + s^-,)}$$

$$= \frac{1}{x_k / (\sum_{j=1}^{n} \lambda_j x_j + s^-,)}$$

$$= \frac{\sum_{j=1}^{n} \lambda_j x_j + s^-}{x_k} = \theta \quad (2.12)$$

The above derivation helps to clarify the nature of the technical efficiency score obtained from the CCR model, namely that the technical efficiency score θ* obtained from the input-oriented CCR model is the ratio of the productivity of the evaluated DMU to the productivity of the benchmark.

2.2.3 Programming Formulation of the Output-Oriented CCR Model

The programming formulation of the output-oriented CCR model is

$$\min \sum_{i=1}^{m} v_i x_{ik}$$

$$s.t. \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0$$

$$\sum_{r=1}^{q} \mu_r y_{rk} = 1$$

$$v \ge 0; \mu \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.13)$$

Its dual model is

$$\max \phi$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge \phi y_{rk}$$

$$\lambda \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.14)$$

The dual model(2.14) measures inefficiency by the extent to which the various outputs can be increased proportionally while the inputs are held fixed, and is therefore called the output-oriented CCR model.

The optimal solution of the model is φ*. At the current technology level, the maximum proportion by which the outputs of the evaluated DMUk can be increased without increasing its inputs is φ*-1. The larger φ* is, the greater the extent to which outputs can be increased and the lower the efficiency. Since φ*≥1, the efficiency score is generally expressed as 1/φ*.[i]

[i]In the analysis results of MaxDEA, the efficiency score (Efficiency Score) of the output-oriented model equals 1/φ*.

2.2.4 The Nature of the Technical Efficiency Score Obtained from the Output-Oriented CCR Model

The model(2.14) can be equivalently expressed as

$$\max \phi$$

$$s.t. \sum_{j=1}^{n}\lambda_j x_{ij} + s_i^- = x_{ik}$$

$$\sum_{j=1}^{n}\lambda_j y_{rj} - s_r^+ = \phi y_{rk}$$

$$\lambda, s^-, s^+ \geq 0 \quad (2.15)$$

In this model too, the benchmark can be understood as $\left(x^*, y^*\right) = (\sum_{j=1}^{n}\lambda_j x_{ij} + s_i^-, \sum_{j=1}^{n}\lambda_j y_{rj} - s_r^+)$, in which case the ratio of the output/input ratio of the evaluated DMU to the output/input ratio of the benchmark (that is, the ratio of the two ratios) is

$$\frac{y_k / x_k}{y^* / x^*} = \frac{y_k / y^*}{x_k / x^*} = \frac{y_k / (\sum_{j=1}^{n}\lambda_j y_j - s^+)}{x_k / (\sum_{j=1}^{n}\lambda_j x_j + s^-,)}$$

$$= \frac{y_k / (\sum_{j=1}^{n}\lambda_j y_j - s^+)}{1}$$

$$= y_k / (\sum_{j=1}^{n}\lambda_j y_j - s^+) = 1/\phi \quad (2.16)$$

It can be seen from the above derivation that the technical efficiency score $1/\phi^*$ obtained from the output-oriented CCR model is likewise the ratio of the productivity of the evaluated DMU to the productivity of the benchmark.

2.2.5 Graphical Illustration of the Input-Oriented CCR Model

The basic principle of the input-oriented CCR model is illustrated graphically below. Suppose there are 7 DMUs (Table 2-3), each with 2 inputs (x1, x2) and 1 output (y). Taking the quantity of input x1 consumed per unit of output as the abscissa (x1/y) and the quantity of input x2 consumed per unit of output as the ordinate (x2/y), the input per unit of output of each DMU is depicted in Figure 2-1 for illustration.

Table 2-3 Example Data for the Input-Oriented CCR Model

DMU	x_1	x_2	y	x_1/y	x_2/y
A	10	40	10	1.00	4.00
B	15	25	10	1.50	2.50
C	32	24	16	2.00	1.50
D	48	16	16	3.00	1.00
E	24	48	16	1.50	3.00
F	54	27	18	3.00	1.50
G	50	60	20	2.50	3.00

In Figure 2-1, each DMU can be represented by a point in the coordinate system. From the coordinate point of the evaluated DMU, perpendiculars are drawn to the two coordinate axes; within the region bounded by the two perpendiculars and the axes (boundary included), the coordinate values of any point are less than or equal to those of the evaluated DMU. Taking G as an example, the coordinate values of all points in the shaded region GQOP, denoted by the point M(m1, m2), are less than or equal to the coordinates of G, denoted G(g1, g2), that is, m1 <= g1 and m2 <= g2. Compared with G, the DMU represented by a point in this region consumes no more of the two inputs per unit of output than G does (the two are equal only when M coincides with G). This means that, apart from G itself, all points in this region have higher technical efficiency than G; it can be seen from the figure that, among the 7 DMUs, the three points B, C and E are more efficient than G.

Among the 7 DMUs, there are 4 DMUs (A, B, C, D) whose perpendiculars, together with the axes, enclose a region that contains no other DMU and no linear combination of other DMUs, which indicates that A, B, C and D lie on the frontier of technical efficiency. The curve formed by joining these four points and its extensions are called the efficient frontier; DMUs lying on the frontier have an efficiency of 1, while DMUs enclosed by the frontier have efficiencies between 0 and 1. In shape, the frontier is convex to the origin and resembles an envelope that encloses all DMUs. In a multidimensional space, the frontier of the input-oriented CCR model is the upper surface of a convex polyhedron that is convex to the origin.

This is the reason why this analytical method is called data envelopment analysis (DEA).

Taking G as an example, the intersection of the line joining G(g1, g2) to the coordinate origin O(0, 0) with the frontier curve is denoted G'(g'1, g'2), and G' is called the projection of G on the frontier. Compared with the point G' on the frontier curve, G consumes g1 - g'1 and g2 - g'2 more of the two inputs per unit of output, respectively; the excess proportions are (g1 - g'1)/g1 and (g2 - g'2)/g2, and the efficient proportions are 1 - (g1 - g'1)/g1 = g'1/g1 and 1 - (g2 - g'2)/g2 = g'2/g2. Expressed in terms of the line segments in the coordinate system, these two proportions are RG'/PG and SG'/QG, respectively, and by geometrical theorems RG'/PG = SG'/QG = OG'/OG. The efficiency score of G can be expressed as OG'/OG. The efficient part of G is represented by OG', and the inefficient part by GG'.

It can be seen from the graphical illustration of the basic principle of DEA that the efficiency measured by DEA is a relative efficiency, namely the efficiency of the evaluated DMU relative to the

"leading" DMUs, and that the efficiency score obtained by DEA results from a comparison of the evaluated DMU with the projection point on the frontier.

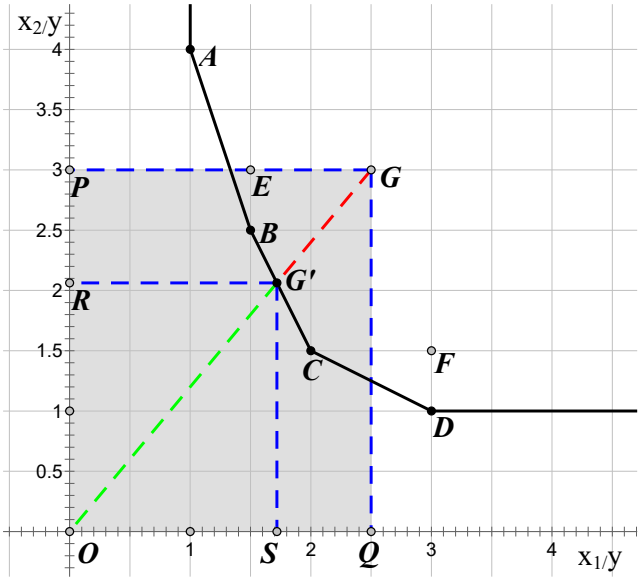


Figure 2-1 Graphical Illustration of the Basic Principle of the Input-Oriented CCR Model

2.2.6 Graphical Illustration of the Output-Oriented CCR Model

Suppose there are 7 DMUs, each with 1 input and 2 outputs (y1 and y2). Taking the quantity of output y1 produced per unit of input as the abscissa (y1/x) and the quantity of output y2 produced per unit of input as the ordinate (y2/x), the output per unit of input of each DMU is depicted in Figure 2-2 for illustration.

Table 2-4 Example Data for the Output-Oriented CCR Model

DMU	x	y_1	y_2	y_1/x	y_2/x
A	10	10	40	1.00	4.00
B	10	25	35	2.50	3.50
C	16	48	40	3.00	2.50
D	16	64	16	4.00	1.00
E	16	32	48	2.00	3.00
F	18	54	27	3.00	1.50
G	20	20	40	1.00	2.00

Taking G as an example, the intersection of the extension of the line joining the coordinate origin O to G with the frontier curve is denoted G', and G' is called the projection of G on the frontier. Per unit of input consumed, the ratios of the output at G' to the output at G are RG'/PG and SG'/QG, respectively, and by geometrical theorems RG'/PG = SG'/QG = OG'/OG. The efficiency status of G can be expressed as φ = OG'/OG. Since φ lies in the range $[1, \infty)$, the reciprocal of φ is generally adopted in practice as the efficiency score, $\theta = 1/\varphi$ = OG/OG', with θ in the range $(0, 1]$.

In a multidimensional space, the frontier of the output-oriented CCR model is the upper surface of a

convex polyhedron that is concave to the origin.

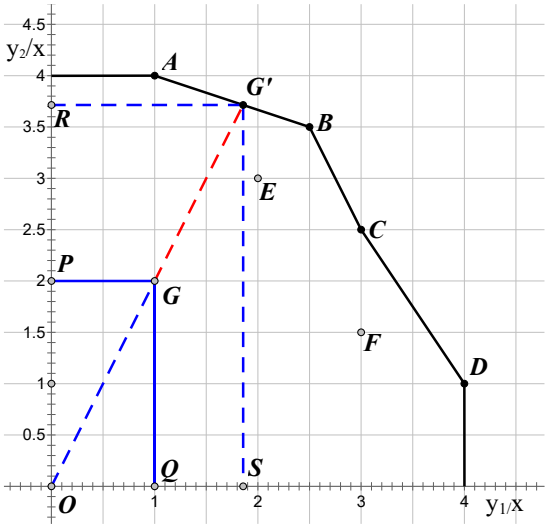


Figure 2-2 Graphical Illustration of the Basic Principle of the Output-Oriented CCR Model

2.3 The BCC Model under Variable Returns to Scale

The CCR model assumes that the returns to scale of the production technology are constant, or that, although the returns to scale of the production technology are variable, all evaluated DMUs are at the stage of optimal production scale, that is, at the constant-returns-to-scale stage. In actual production, however, many production units are not in the optimal scale of production, so the technical efficiency obtained from the CCR model contains a component of scale efficiency. In 1984, Banker, Charnes and Cooper published "Some models for estimating technical and scale inefficiencies in data envelopment analysis" in Management Science, and proposed a DEA model for estimating scale efficiency(Banker, Charnes, & Cooper, 1984). The proposal of this method is of great significance for the DEA theoretical method, and in the subsequent literature this model is called the BCC model (named after the initial letters of the surnames of the three authors). The BCC model is based on variable returns to scale (VRS), and the technical efficiency it yields excludes the influence of scale and is therefore called "pure technical efficiency" (PTE).

2.3.1 Programming Formulation of the Input-Oriented BCC Model

The BCC model is obtained from the CCR dual model(2.10) by adding a convexity constraint $\sum_{j=1}^{n}\lambda_j=1$, which serves to construct, on the basis of a variable-returns-to-scale technology, a pure technical efficiency frontier from which scale efficiency has been removed, so that the comparison between the evaluated DMU and its projection point is made at the same returns to scale (see the graphical illustration below).

$$\min \theta$$

$$s.t. \sum_{j=1}^{n}\lambda_j x_{ij} \le \theta x_{ik}$$

$$\sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{rk}$$

$$\sum_{j=1}^{n}\lambda_j=1$$

$$\lambda \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n \quad (2.17)$$

The dual programming formulation of the BCC model(2.17) is[i]

[i] μ0 is a free variable taking values in (-∞, +∞), so whether the sign in front of μ0 in the linear programming formulation is positive or negative does not affect the value of the objective function. Consistent with the original paper(Banker et al., 1984) a negative sign is adopted in this book.

$$\max \sum_{r=1}^{s} \mu_r y_{rk} - \mu_0$$

$$s.t. \sum_{r=1}^{q} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - \mu_0 \le 0$$

$$\sum_{i=1}^{m} v_i x_{ik} = 1$$

$$v \ge 0; \mu \ge 0; \mu_0 \ free$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.18)$$

2.3.2 The Nature of the Technical Efficiency Score Obtained from the Input-Oriented BCC Model

The technical efficiency score θ^* obtained from the input-oriented BCC model is likewise the ratio of the productivity of the evaluated DMU to the productivity of the benchmark. The difference from the input-oriented CCR model is that the benchmark of the BCC model lies on the pure technical efficiency frontier.

2.3.3 Programming Formulation of the Output-Oriented BCC Model

The programming formulation of the output-oriented BCC model is

$$\max \phi$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge \phi y_{rk}$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.19)$$

The output-oriented BCC model is obtained from the output-oriented CCR model(2.14) by adding a convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$.

Its dual programming formulation is

$$\min \sum_{i=1}^{m} v_i x_{ik} + v_0$$

$$s.t. \sum_{r=1}^{q} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - v_0 \le 0$$

$$\sum_{r=1}^{s} \mu_r y_{rk} = 1$$

$$v \ge 0; \mu \ge 0; v_0 \ free$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.20)$$

2.3.4 The Nature of the Technical Efficiency Score Obtained from the Output-Oriented BCC Model

The technical efficiency score $1/\phi^*$ obtained from the output-oriented BCC model is likewise the ratio of the productivity of the evaluated DMU to the productivity of the benchmark. The difference from the output-oriented CCR model is that the benchmark of the BCC model lies on the pure technical efficiency frontier.

2.3.5 Graphical Illustration of the Input-Oriented BCC Model

The basic principle of the input-oriented BCC model is illustrated graphically below. Suppose there are 4 DMUs (Table 2-5), each with 1 input (x) and 1 output (y). With input x as the abscissa and output y as the ordinate, the input-output situation of each DMU is shown in the figure (Figure 2-3).

Table 2-5 Example Data for the Input-Oriented BCC Model

DMU	x	y
A	1.00	1.00
B	1.50	2.00
C	3.50	2.50
D	4.00	3.00

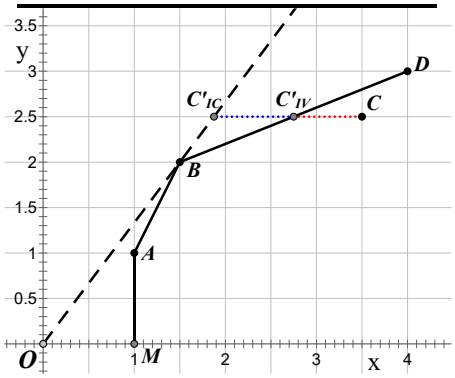


Figure 2-3 Graphical Illustration of the Basic Principle of the Input-Oriented BCC Model

Suppose that returns to scale are constant (CRS); the production frontier can then be represented by the ray OB, and B is the only efficient DMU. Suppose that returns to scale are variable (VRS); the production frontier is then the curve MABD, which is convex to the left. In the CRS model, point B has the highest output/input ratio, and points A, C and D are all inefficient relative to B: the projection of A lies on OB with a λ value of less than 1, whereas the projections of C and D lie on the extension of OB with λ values greater than 1. In the VRS model, however, restricted by the convexity constraint $\sum_{j=1}^{n}\lambda_j=1$, none of the three points A, C and D can be projected onto the ray OB, because their projection points on the ray OB either $\sum_{j=1}^{n}\lambda_j<1$, or $\sum_{j=1}^{n}\lambda_j>1$, and therefore cannot satisfy $\sum_{j=1}^{n}\lambda_j=1$. The convexity constraint $\sum_{j=1}^{n}\lambda_j=1$ causes the production frontier to shrink back and places points A and D on the frontier; ABD together with the vertical downward extension through A constitute the "pure technical efficiency" production frontier curve. In the VRS model, point C is an inefficient DMU, and its input inefficiency is represented by CC'IV, whereas in the CRS model its input inefficiency is represented by CC'IC.

2.3.6 Graphical Illustration of the Output-Oriented BCC Model

The basic principle of the output-oriented BCC model is now illustrated graphically. Again taking Table 2-5 as an example, in the CRS model the output-oriented and the input-oriented frontiers are the same, both being the ray OB. In the VRS model, however, the output-oriented frontier is not entirely the same as the input-oriented frontier: the output-oriented "pure technical efficiency" production frontier curve is ABD together with the extension through D parallel to the x axis (Figure 2-4). The projection of point C on the frontier is entirely different: its inefficiency is represented by CC'OV, whereas in the CRS model its output inefficiency is represented by CC'OC.

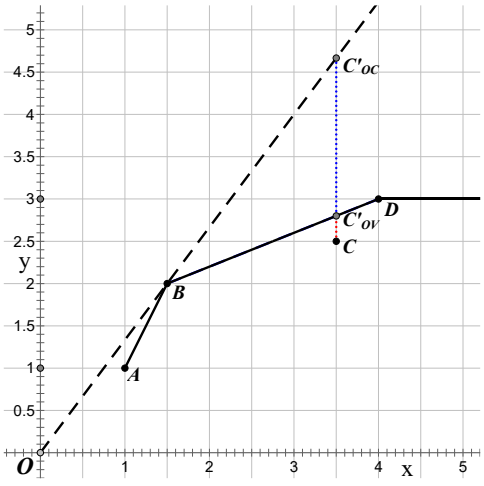


Figure 2-4 Graphical Illustration of the Basic Principle of the Output-Oriented BCC Model

2.4 The Comparability of Efficiency Scores among DMUs

The technical efficiency score obtained from a DEA model is the ratio of the productivity of the evaluated DMU to the productivity of the benchmark. If all DMUs referred to the same benchmark, there would be no doubt that the efficiency scores of the various DMUs would be comparable with one another. In practice, however, the benchmarks of different DMUs are different.

Efficiency scores remain comparable across DMUs by virtue of the equivalence of the production frontier in terms of technical efficiency, that is, all points on the production frontier (each representing a virtual DMU) have a technical efficiency of 100%. Although different DMUs have different benchmarks, the technical efficiency of all benchmarks is the same.

The production frontier is the core of the DEA model and plays a key role in the measurement of efficiency scores. A stable production frontier is the guarantee of reliable results. Constructing a stable production frontier requires a sufficient number of DMUs.

2.5 The Relationship between Input-Oriented and Output-Oriented Efficiency Scores

In the CRS model, the efficiency scores obtained under input orientation and under output orientation are equal, which can be illustrated graphically (Figure 2-5). The efficiency score of point C can be expressed as QC'IC/QC in the input-oriented model (θ*) and as NC/NC'OC in the output-oriented model (1/φ*). By the theorems of similar triangles it is easy to prove that C'ICC/QC = CC'OC/NC'OC, from which QC'IC/QC = NC/NC'OC follows. In the VRS model, however, no such relationship holds; that is, in the VRS model the input-oriented and the output-oriented efficiency scores are not equal.

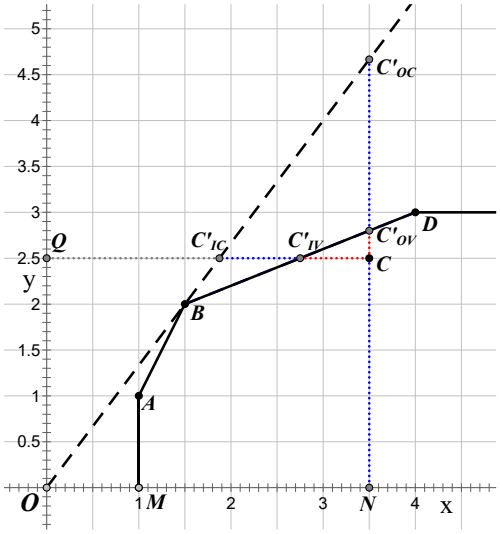


Figure 2-5 Graphical Illustration of the Equality of Input-Oriented and Output-Oriented Efficiency Scores in the CRS Model

2.6 The Scale Efficiency Problem

The BCC model was proposed in order to solve for the technical efficiency of DMUs under a VRS production technology, but it also provides a method for calculating scale efficiency. If the production technology exhibits variable returns to scale, the efficiency score obtained from the CRS model (technical efficiency, TE) is not a pure technical efficiency but contains a scale efficiency component, which makes it possible to solve for scale efficiency. For a VRS production technology, since the efficiency score obtained from the VRS model is the technical efficiency (called "pure technical efficiency", PTE), comparing the calculated CRS efficiency score with the VRS efficiency score allows the scale efficiency (SE) to be isolated; the calculation is SE = TE/PTE.

In Figure 2-6 the scale inefficiency of the input-oriented model is expressed as C'IVC'IC, and in Figure 2-7 the scale inefficiency of the output-oriented model is expressed as C'OCC'OV.

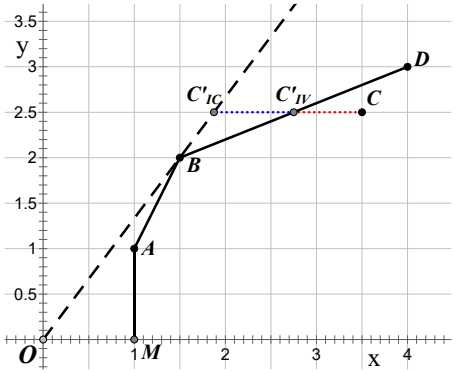


Figure 2-6 Illustration of Scale Efficiency in the Input-Oriented Model

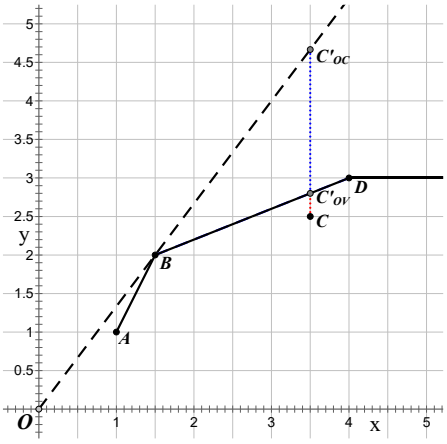


Figure 2-7 Illustration of Scale Efficiency in the Output-Oriented Model

It can be seen from the above figures that the scale efficiency of a DMU obtained from a DEA model is not the scale efficiency of that DMU itself but the scale efficiency of the projection point of that DMU on the frontier, calculated by means of that DMU. The frontier projection points of the input-oriented and output-oriented models differ, and the scale efficiencies obtained therefore also differ.

It should be noted that a scale efficiency problem exists only when the production technology exhibits variable returns to scale (VRS). In solving for scale efficiency, the CRS model is merely borrowed

(assuming that the DMUs under study are in a CRS state) in order to isolate scale efficiency. The logical relationship involved must be clearly understood.

According to the definition of scale efficiency, a radial distance must be used to calculate scale efficiency. When other types of distance are used[i], the resulting TE/PTE ratio is not "scale efficiency" (Scale Efficiency); in MaxDEA this similar ratio is called "scale effect" (Scale Effect), in order to distinguish between the two.

William W Cooper, Seiford, & Tone (2007) called the difference between the SBM efficiency score and the radial efficiency score MIX (Mix Efficiency), that is, SBM efficiency score = radial efficiency score × MIX. Under this definition, the following relationship can be derived

$$\text{SBM Scale Effect} = \frac{\text{SBM CRS Efficiency}}{\text{SBM VRS Efficiency}} = \frac{\text{Radial CRS Efficiency} \times \text{CRS MIX}}{\text{Radial VRS Efficiency} \times \text{VRS MIX}}.$$
$$= \text{Scale Efficiency} \times \text{Scale MIX}$$

[i]Other distance types are described in detail in Chapter 4.

2.7 The Naming of DEA Models

The CCR model marks the birth of the DEA method, and the proposal of the BCC model is likewise of very great significance for the development of the DEA method. Naming these two models after the surnames of their authors has therefore met with unanimous acceptance in the DEA field. However, with the rapid development of DEA theory and methods, a large number of new DEA methods have continually appeared, and if models continue to be named after the surnames of their authors, confusion is easily caused. DEA models should preferably be named according to their main characteristics. Under such a naming convention, the CCR model is a CRS radial DEA model and the BCC model is a VRS radial DEA model. The term radial means that inefficiency is measured by the extent to which inputs can be reduced proportionally, or the extent to which outputs can be increased proportionally. In order to facilitate the presentation of more DEA models, the two names CCR and BCC will no longer be used in the later chapters of this book.

2.8 Strong Efficiency, Weak Efficiency and the Slack Problem

Why are there slack variables in DEA models? In the radial DEA envelopment model, the linear programming formulations(2.10) (2.14)express the constraints in the form of inequalities rather than equalities. This can be regarded as a kind of "loose" constraint, and it is the basis for the existence of slack variables.[i]

The following everyday problem can help in understanding slack variables in radial DEA models: there are several ropes of different lengths; one end of each rope is fixed to a wall, and the other ends of all the ropes are then aligned and pulled outward. When the ropes can be pulled no further, the shortest rope has become taut and cannot be pulled any further. At this point, however, the other ropes may still be slack; if the taut rope is released, the slack ropes can still be pulled further, and the distance by which each slack rope can be pulled further is the slack variable value of the corresponding measure.

The reason why the slack problem arises can also be understood in terms of the linear space of the envelopment model: the frontier of a DEA model is formed by a piecewise linear function, and a piecewise linear function may contain segments parallel to the coordinate axes in the spatial coordinate system; this is the root cause of the slack problem.

Consider first the example of a single input and a single output, again taking Table 2-5 as an example. In the CRS model, the frontier for a single input and a single output is the ray OB, and segments parallel to the coordinate axes cannot occur; the slack variables of the CRS radial model with a single input and a single output are therefore certainly 0, that is, no slack problem exists. The VRS frontier, however, does contain segments parallel to the coordinate axes, and if the projection of a DMU falls within such a parallel segment, a slack problem exists. For example, on the basis of the data in Table 2-5 a point S (x=2, y=0.5) is added. In the input-oriented VRS radial model, the projection of point S on the frontier is S', but compared with point A on the frontier, S' exhibits insufficient output (the segment S'A), that is, a slack problem exists (Figure 2-8).

[i] In the theory of production possibility sets, the production possibility set of a DEA model is expressed as $\left\{(\mathrm{x},\mathrm{y}):\mathrm{x}\geq\sum_{j=1}^{n}\lambda_j x_{ij},\ \mathrm{y}\leq\sum_{j=1}^{n}\lambda_j y_{rj};i=1,2\ldots,m;r=1,2\ldots,q;\ j=1,2\ldots,n\right\}$ Such a "loose" constraint using inequality signs represents the free disposability of inputs and outputs, which means: if input x can produce output y, then a larger input x+ Δ x can also produce y; and if x can produce y, then x can also produce a smaller output y- Δ y.

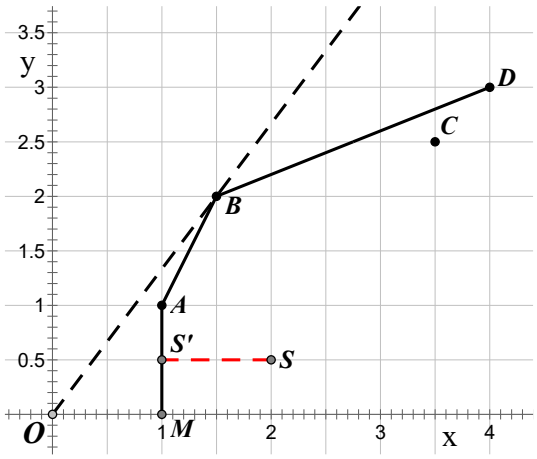


Figure 2-8 Illustration of the Slack Problem in the VRS Model with a Single Input and a Single Output

In the case of multiple inputs and multiple outputs, a slack problem arises in both the CRS model and the VRS model.

Figure 2-9 shows the problem of slack variables in an input-oriented CRS model with two inputs and one output. The projection of point E on the frontier falls on E', which lies in a segment parallel to the coordinate axes, and the slack in its input per unit of output, x2/y, is E'B.

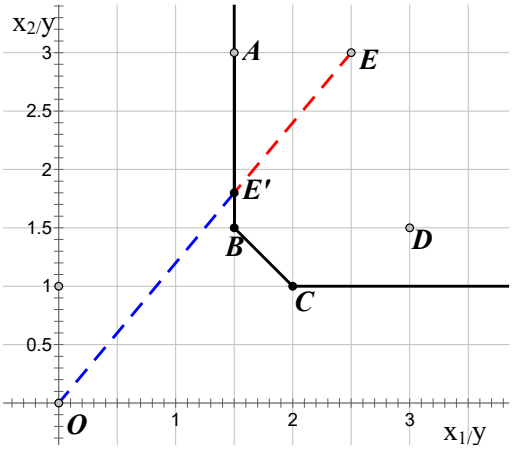


Figure 2-9 Illustration of the Slack Problem in the CRS Model with Two Inputs and One Output

In DEA models, when the number of DMUs is too small, the frontier surface constructed is insufficiently refined, and segments parallel to the coordinate axes are more likely to appear within the projection region of the evaluated DMU, so that a slack problem arises.

Slack variable values can be solved for in several ways. Taking the input-oriented model as an example, if slack variables are added only to the constraints of the model and not to the objective function, that is,

$$\min \theta$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rk}$$

$$\sum_{j=1}^{n}\lambda_j=1 \quad (2.21)$$

This approach is usually called the one-stage method. In most cases the one-stage method generally yields correct slack variable values; take Figure 2-8 as an example. However, the one-stage method cannot guarantee that complete slack variable values are always obtained; for example, in Figure 2-9 two DMUs lie within the segment parallel to the coordinate axes in which the projection point of E lies, and in this case the one-stage method cannot solve for the slack variable of input x2/y. When solving DEA models, if the user is concerned only with the efficiency score, the one-stage method may be chosen.

To remedy the incompleteness of the slack variable values obtained by the one-stage method, the input and output slack variables s- and s+ must be added to the objective function of the model. Taking the input-oriented radial model as an example, taking the model(2.21) as the basis, the programming formulation obtained by adding the slack variables to the objective function is:

$$\min \theta-\varepsilon\sum(s^{-}+s^{+})$$

$$s.t.\ \sum_{j=1}^{n}\lambda_j x_{ij}+s_i^{-}=\theta x_{ik}$$

$$\sum_{j=1}^{n}\lambda_j y_{rj}-s_r^{+}=y_{rk}$$

$$\sum_{j=1}^{n}\lambda_j=1$$

$$\lambda\geq 0;\ s^{-}\geq 0; s^{+}\geq 0;$$

$$i=1,2\ldots,m; r=1,2\ldots,q;\ j=1,2\ldots,n \quad (2.22)$$

ε is a constant denoting the non-Archimedean infinitesimal, which may be understood as a sufficiently small positive number, for example 0.00001 in practical applications. Improper setting of ε gives rise to computational errors, and if ε is set too large it may even make the objective function unbounded.

The multiplier form (dual model) of the model(2.22) is

$$\max \sum_{r=1}^{s} \mu_r y_{rk} - \mu_0$$

$$s.t. \sum_{r=1}^{q} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - \mu_0 \le 0$$

$$\sum_{i=1}^{m} v_i x_{ik} = 1$$

$$v \ge \varepsilon$$

$$\mu \ge \varepsilon$$

$$\mu_0 \ free$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.23)$$

If ε is set too large, the model may be left without a feasible solution.[i]

To eliminate the error caused by ε in the envelopment model(2.22), the two-stage method may be used to solve for the input and output slack variables(Ali & Seiford, 1993).

In the first stage, the model(2.17) is solved; then in the second stage, the following programming formulation is solved:

$$\max \sum (s^- + s^+)$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta_k^* x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rk}$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda \ge 0;\ s^- \ge 0; s^+ \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \quad (2.24)$$

θ* denotes the optimal value of θ obtained in the first stage.

In Figure 2-8 the slack variable of output y is 0.5. In Figure 2-9 the slack variable of input x2/y is 0.3.

In the output-oriented model, the slack problem can likewise be solved in a similar manner.

The efficiency status of the evaluated DMU, efficient or inefficient, can be obtained from a DEA model. DEA efficiency takes two forms:

The first is strong efficiency, which refers to a state of production in which the quantity of no input can be reduced unless the quantity of an output is reduced or the quantity of another input is increased, and the quantity of no output can be increased unless the quantity of an input is increased or the quantity of

[i]By duality theory, if the objective function of the primal model is unbounded, then its dual model has no feasible solution.

another output is reduced. Such a production state is a Pareto-optimal state. In radial DEA models, the criterion for strong efficiency is θ* = 1, with s-* and s+* both equal to 0.

The second is weak efficiency, which refers to a state in which the quantities of all inputs cannot be reduced proportionally unless the quantity of output is reduced, and the quantities of all outputs cannot be increased proportionally unless the quantity of input is increased. In this production state, although inputs cannot be reduced proportionally nor outputs increased proportionally, one or several (but not all) inputs may nevertheless be reducible, or one or several (but not all) outputs may nevertheless be increasable; such a production state is therefore called weakly efficient. In radial DEA models, the criterion for weak efficiency is whether θ* equals 1, without regard to whether a slack problem exists.

Whether a DMU is efficient and whether slacks exist are two different questions. In radial DEA models a slack problem does not necessarily exist, and it is also possible for all slack variable results to be 0.

The first case is the CRS model with only one input and one output. The frontier of the single-input, single-output CRS model is a point, so the problem of segments parallel to the coordinate axes does not arise, and hence no slack problem exists either.

The second case is that the projection points of all inefficient DMUs on the frontier lie in the strongly efficient part of the frontier; in this case all slack variable results for all inefficient DMUs are likewise all 0, and no slack problem arises.

2.9 The Relationship between Radial (Proportional) Improvement and Slack Improvement

The projection point of an inefficient DMU on the frontier represents its target value. If input improvement values are expressed as negative numbers and output improvement values as positive numbers[i], then the projected input and output values of the evaluated DMU can be expressed uniformly as:

Projected value = original value + improvement value.

The improvement value of an inefficient DMU comprises two parts: the proportionate movement[ii] and the slack movement. That is,

Projected value = original value + proportionate improvement value + slack improvement value.

If an inefficient DMU carries out only the proportionate improvement, it may become weakly efficient after the improvement: if a slack problem exists, it is weakly efficient; if no slack problem exists, it is strongly efficient. Only after both the proportionate improvement and the slack improvement have been carried out is the inefficient DMU guaranteed to become strongly efficient.

Taking Figure 2-10 as an example, D is an inefficient DMU; after the proportionate improvement it moves from point D to point C on the frontier, and since point C lies in the strongly efficient part of the frontier, where no slack problem exists, D is strongly efficient once the proportionate improvement has been made. E is also an inefficient DMU; after the proportionate improvement it moves from point E to point E' on the frontier, and since point E' lies in the weakly efficient part of the frontier (the part parallel to the coordinate axes), where a slack problem exists, E is only weakly efficient after the proportionate improvement; only when the slack improvement is then made, moving from point E' to point B, does it become strongly efficient.

[i]Note that in super-efficiency models the sign of the "improvement value" of a super-efficient DMU is the opposite of that in the standard efficiency model, and the "improvement value" has a different meaning in super-efficiency models; see "Chapter 5 Super-efficiency Models" for details.

[ii]In radial models the proportional improvement is generally called radial movement, denoting the amount by which each input (or output) is improved in the same proportion. In directional distance function models, however, this part of the improvement is not a proportional improvement but follows the direction of the direction vector. To unify the terminology for this component of improvement, it is referred to in the analysis results of MaxDEA simply as the proportional improvement.

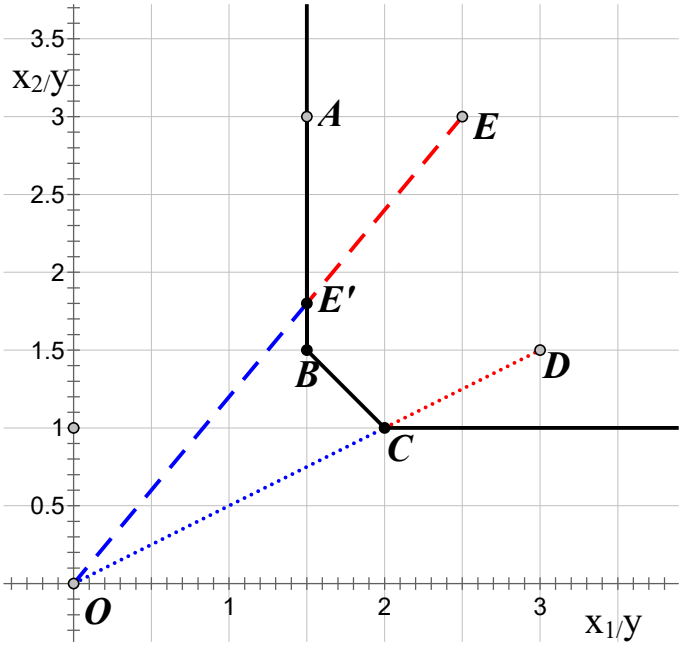


Figure 2-10 Graphical Illustration of the Slack Problem in the CRS Model with Two Inputs and One Output

2.10 Choosing the Number of DMUs, the Input–Output Measures and the Model Orientation

2.10.1 The Requirement of DEA Models on the Number of DMUs

DEA is a non-parametric frontier analysis method. Although DEA imposes relatively few requirements on the number of DMUs compared with parametric methods, if the number of DMUs is too small—for example, if the number of DMUs is smaller than the number of input and output measures (n<m+q)—it is very easy to obtain a result in which most or even all DMUs are efficient, so that DEA loses its ability to discriminate among the efficiencies of DMUs. Generally speaking, the number of DMUs should be no less than the product of the numbers of input and output measures, and at the same time no less than three times the number of input and output measures(William W Cooper et al., 2007), that is,

$$n \geq \max\{m \times q, 3 \times (m + q)\}.$$

This, however, is only a rough guiding principle, and the specific situation must be judged from the DEA analysis results. In practical applications the number of DMUs is often fixed, and when the discriminating power of the model is insufficient, the only option is to reduce the number of input or output measures in order to improve discrimination.

When the number of DMUs is fixed, the greater the number of input and output measures included in the model, the larger the efficiency score obtained. If one input or output measure is added to the original model (whether the CCR or the BCC model), the efficiency score obtained from the new model is always greater than or equal to the efficiency score obtained from the original model.

2.10.2 Choosing the Input and Output Measures

From the perspective of the production possibility set, the input measures (x) and output measures (y) of a DEA model should broadly satisfy the following relationships:

1) x can produce y;

2) y is produced from x.

In practical applications, care should be taken to distinguish input and output measures from the factors that influence efficiency (or productivity). Taking hospitals as an example, their main output measures are outpatient and emergency services and inpatient services, and their main input measures include health technical personnel, beds and other medical equipment and facilities. The reform of the pharmaceutical and healthcare system currently under way will affect the efficiency (or productivity) of hospitals, but the reform is an external factor influencing efficiency (or productivity), not an input measure.

DEA does not require the absence of high correlation (collinearity) among input measures or among output measures, and the existence of collinearity does not lead to erroneous analysis results. In fact, in many production practices two (or more) factors of production are objectively required to be employed in a certain proportion. Taking hospitals as an example, there is a reasonable range for the ratio of doctors to nurses, and in actual data the number of doctors and the number of nurses are highly correlated.

However, when a large number of measures results in insufficient discriminating power of the model, then, from the standpoint of minimizing the number of input and output measures while including as many factors of production as possible, the highly correlated measures may be considered first for exclusion from the model.

The DEA model is a linear programming method based on production possibility set theory; the production possibility set is a linear combination of DMUs, and in theory the input and output measures in a DEA model must be capable of linear addition. A rate (or ratio) is obtained by dividing a numerator measure by a denominator measure, and if the denominator values of the rate measure differ across DMUs (in practice they usually do), an erroneous production possibility set will be generated and illogical results may also arise (Emrouznejad & Amin, 2009).(Emrouznejad & Amin, 2009) For example, many of the measures reflecting the inputs and outputs of the health system take the form of rates or ratios, such as total health expenditure as a proportion of GDP, service coverage, mortality, incidence and prevalence. If rates are used directly as input or output measures, the results of projection analysis in the CRS model may include projected rate values greater than 100%. Although the use of a VRS model avoids the problem of projected rate values exceeding 100%(Hollingsworth & Smith, 2003), the VRS model still suffers from the problem that the production frontier determined by the DEA model lies outside the production possibility set(Emrouznejad & Amin, 2009).

In a VRS model, when the denominator values of a rate measure differ little across DMUs, the rate measure may be regarded approximately as satisfying the condition of linear addition; alternatively, when the error caused by the failure of linear addition is acceptable for the research purpose, the rate measure may be used as an input or output measure.

Emrouznejad & Amin (2009) suggested that when a rate is present among the input or output

measures, two methods can be used to handle such measures.

Taking the output-oriented VRS model as an example, the first method takes the numerator of the rate as an output measure and the denominator of the rate as an input measure, and its linear programming formulation is

$$\begin{aligned}
&\max\ \phi \\
&s.t.\ X\lambda \le x_0 \\
&\quad Y\lambda \ge \phi y_0 \\
&\quad N\lambda \ge \phi n_0 \\
&\quad D\lambda \le d_0 \\
&\quad \sum_j \lambda_j = 1 \\
&\quad \lambda \ge 0
\end{aligned} \tag{2.25}$$

where N and D denote the numerator and the denominator of the rate measure respectively, and n0 and d0 denote the numerator and the denominator of this rate measure for the evaluated DMU respectively.

The second method splits the rate measure into a numerator measure and a denominator measure on the left-hand side of the linear programming formulation, carrying out the linear operations separately, but does not split the rate measure of the evaluated DMU on the right-hand side of the equation; its programming formulation is

$$\begin{aligned}
&\max\ \phi \\
&s.t.\ X\lambda \le x_0 \\
&\quad Y\lambda \ge \phi y_0 \\
&\quad \frac{N\lambda}{D\lambda} \ge \phi r_0 \\
&\quad \sum_j \lambda_j = 1 \\
&\quad \lambda \ge 0
\end{aligned} \tag{2.26}$$

where N and D denote the numerator and the denominator of the rate measure respectively, and r0 denotes the rate measure of the evaluated DMU.

The advantage of the first method is that the linear programming form of the model is preserved; its disadvantage is that the number of input or output measures is increased, which raises the requirement on the sample size. The advantage of the second method is that the number of input and output measures is not increased, but the model becomes a nonlinear programming problem. Both methods presuppose that the numerator and denominator data of the rate measure can be obtained.

2.10.3 Choosing the Model Orientation

According to the way in which efficiency is measured, DEA models can be divided into input-oriented, output-oriented and non-oriented models (the last to be introduced later). The input-oriented model measures the degree of inefficiency of the evaluated DMU from the input perspective, focusing on the extent to which each input should be reduced in order to attain technical efficiency without reducing outputs; the output-oriented model measures the degree of inefficiency of the evaluated DMU from the output perspective, focusing on the extent to which each output should be increased in order to attain technical efficiency without increasing inputs; the non-oriented model measures from the input and output sides simultaneously.

With specific reference to the radial models introduced in this chapter, input orientation means that the state of inefficiency is measured by the extent to which each input can be reduced proportionally given the outputs, while output orientation means that the state of inefficiency is measured by the extent to which each output can be increased proportionally given the inputs. In DEA models, "given the outputs" in practice means "without reducing the outputs", and "given the inputs" in practice means "without increasing the inputs".

The choice of model orientation depends mainly on the purpose of the analysis. If the purpose is merely to obtain the efficiency scores of the units, any of the three orientations will do. If further projection analysis is required, then from a management perspective, if reducing inputs is regarded as the main route by which inefficient units improve their efficiency, the input-oriented model should be chosen; if increasing outputs is regarded as the main route to higher efficiency, the output-oriented model should be chosen. The choice of orientation must be understood in conjunction with the specific field of the analysis. In the health field, for example, if health resources are underinvested, choosing the input-oriented model makes the results of projection analysis difficult to interpret. This is because in the input-oriented model, according to the results of projection analysis, the improvement target for an inefficient unit to reach an efficient state is to reduce inputs, which appears to contradict the background of insufficient investment in health resources (although in essence it does not), so that the analysis results are easily misunderstood. In addition, when analysing the efficiency of health services under conditions of insufficient demand (when demand is the primary factor determining the quantity of health service outputs), choosing the output-oriented model makes the output targets determined by projection analysis objectively difficult to achieve, so that the analysis loses practical guiding significance(成刚 & 钱振华, 2012).

2.11 A Worked Example: Solving Basic DEA Models in MaxDEA

This section uses actual data to illustrate the process and results of solving basic DEA models with MaxDEA. Table 2-6 presents some of the input and output measures of hospitals in each province (municipality and autonomous region) in 2011. A DEA model is constructed with the number of beds and the number of health technical personnel as input measures and the number of outpatient and emergency visits and the number of admissions as output measures. The basic DEA models comprise four radial DEA models: the input-oriented CRS model, the output-oriented CRS model, the input-oriented VRS model and the output-oriented VRS model. To save space, this section demonstrates only the CRS models in detail.

Table 2-6 Selected Input and Output Measures of Hospitals by Region (2011)

Region	Number of beds	Number of health technical personnel	Number of outpatient and emergency visits	Number of admissions
Anhui	140997	132739	63344221	4391516
Beijing	87596	128644	104340626	1882593
Fujian	89947	93898	72059496	3087618
Gansu	66661	53127	27719369	1710826
Guangdong	246050	289388	293784406	7993062
Guangxi	95752	105773	64449855	3258438
Guizhou	78368	69320	29105344	2433370
Hainan	21367	26329	12692336	611621
Hebei	187504	183683	82489933	5675067
Henan	239793	231149	114076405	7043628
Heilongjiang	129449	127358	47962073	2992970
Hubei	152062	147628	84006816	4805124
Hunan	168428	152904	63238363	5393968
Jilin	94636	86278	38087139	2259264
Jiangsu	221674	214938	166944454	6433837
Jiangxi	87184	93287	46383376	3000964
Liaoning	171032	156893	70270696	4032159
Inner Mongolia	72871	74177	31379522	1747905
Ningxia	22037	21830	12373335	602227
Qinghai	18586	16689	8918596	469811
Shandong	280385	281654	134719233	8354234
Shanxi	110741	116660	37390228	2332714
Shaanxi	114339	123537	52692308	3236715
Shanghai	87548	95198	113669022	2282773
Sichuan	211524	194752	110475915	6489314
Tianjin	40787	53543	52164779	1037054

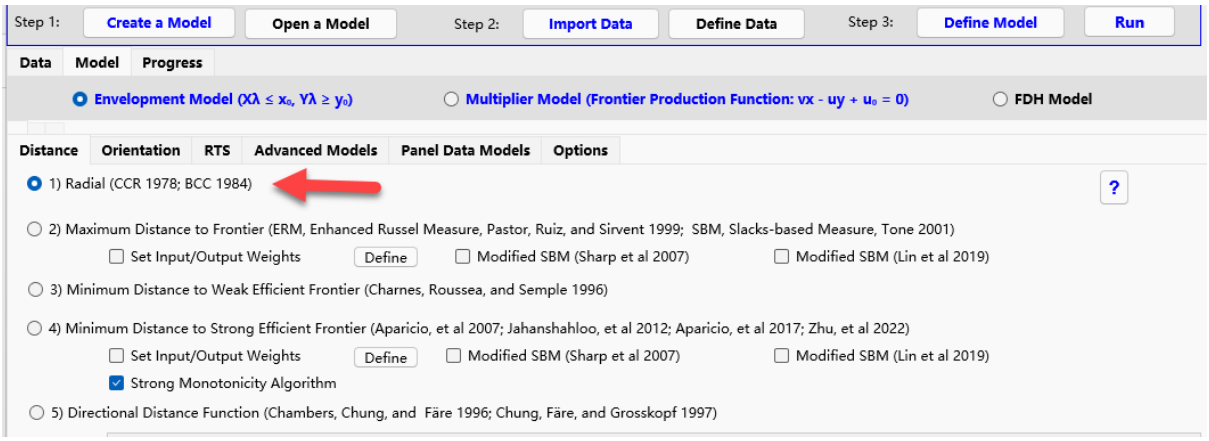
Tibet	6314	5843	3998564	120766
Xinjiang	97436	84202	37383273	3115413
Yunnan	126318	90330	62327579	3736495
Zhejiang	162905	187137	181663463	4899310
Chongqing	74827	66653	38726559	2116631

2.11.1 Model Settings

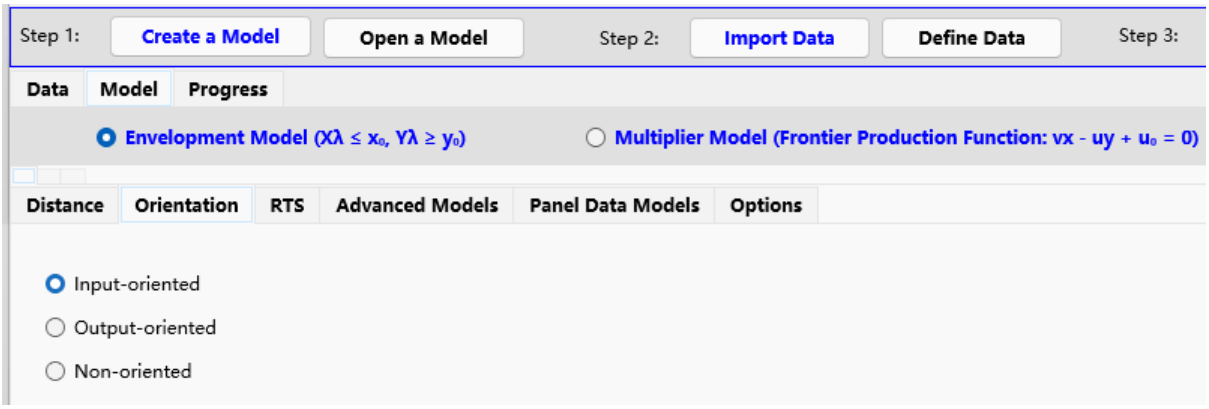
2.11.1.1 Detailed Setup

The basic settings of a radial model comprise three aspects: "Distance" (the distance function type), "Orientation" and "Returns to Scale".

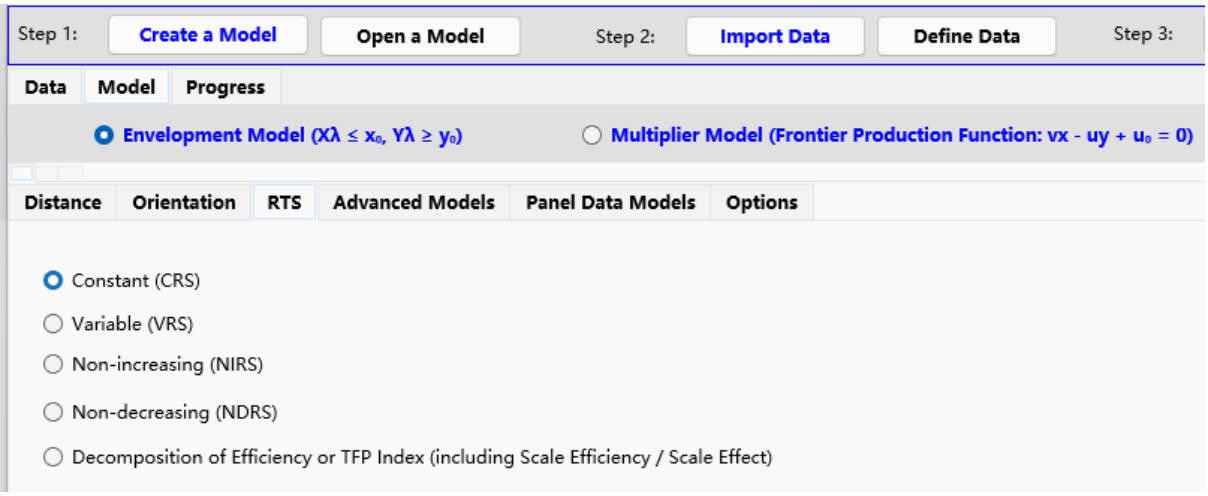
For Distance, select "1) Radial distance (CCR 1978; BCC 1984)".



For Orientation, select "1) Input-oriented" or "2) Output-oriented".



Returns to Scale: select "Constant (CRS)" for the CCR model and "Variable (VRS)" for the BCC model.



2.11.1.2 Shortcut

MaxDEA provides a shortcut menu for the four basic DEA models. Through the menus "Model" and "Run Basic Models in One Click", the four basic DEA models can be run directly: the input-oriented CRS model, the output-oriented CRS model, the input-oriented VRS model and the output-oriented VRS model. Note that what is run through the shortcut is the envelopment model.

2.11.2 Analysis Results

The results of the envelopment model include:

1) Score: efficiency score
2) Rank: ranking of the efficiency score
3) Benchmark and Lambda: the benchmarks and their coefficients
4) Proportionate Movement: the proportionate change (radial movement); a positive number indicates that the projection direction is an increase in quantity, and a negative number indicates a decrease in quantity
5) Slack Movement: the slack variable value; a positive number indicates that the projection direction is an increase in quantity, and a negative number indicates a decrease in quantity
6) Projection: target value (projected value)
7) Dual Price: the shadow prices of the inputs and outputs, corresponding to the input and output weights in the analysis results of the multiplier model[i]. The specific

[i] In the dual solution of the envelopment model the shadow prices of inputs are negative and the shadow prices of outputs are positive; in the multiplier model the weight coefficients of both inputs and outputs are positive. Since multiple optimal solutions may exist, the shadow prices of inputs and outputs in the envelopment model may not coincide with the weights of inputs and outputs in the multiplier model.

meaning of the dual solution is given in textbooks on linear programming or operations research.

The results of the multiplier model include:

1) Score: efficiency score
2) Rank: ranking of the efficiency score
3) Weight: the weight coefficients of the inputs and outputs
4) RTS Constant: the constant term of returns to scale[i]
5) Benchmark and Lambda: the benchmarks and their coefficients
6) Projection: target value (projected value)

2.11.2.1 Efficiency Scores and Slack Variables

The following table reports the efficiency scores and slack variables of the input-oriented CRS and output-oriented CRS models. In the analysis results, the first column, NO, is the serial number, the second column is the name of the DMU and the third column is Score, that is, the efficiency score: in the input-oriented model, Score = θ*, and in the output-oriented model, Score = 1/φ*. Slack variables are reported in the form Slack Movement (measure name), so the slack result for the number of beds, for example, is reported as Slack Movement (number of beds). For an inefficient DMU, the direction of improvement is to reduce inputs or to increase outputs. In the MaxDEA analysis results, therefore, slack improvements in inputs are reported as negative numbers and slack improvements in outputs as positive numbers.

Note that Slack in a DEA model is itself non-negative; in the output of the MaxDEA software the name Slack Movement is used and a sign is attached to the value in order to indicate the direction in which the input or output slack changes (a negative sign indicates a decrease and a positive sign an increase). This point is particularly important in the results of the super-efficiency models in Chapter 5. If the original value of Slack in the model is required, it is simply the absolute value of Slack Movement.

[i]For the CRS model this item is 0, so it need not be displayed in the results of the CRS model.

Table 2-7 Efficiency Scores and Slack Variables of the Input-Oriented CRS and Output-Oriented CRS Models[i]

	Input-oriented CRS					Output-oriented CRS				
DMU	Efficiency score	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions	Efficiency score	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions
Anhui	0.944	0	0	11623213	0	0.944	0	0	12309544	0
Beijing	0.917	0	-30637	0	212841	0.917	0	-33394	0	231997
Fujian	1.000	0	0	0	0	1.000	0	0	0	0

The regularities of the efficiency scores and slack variables, and the features of the results of this example, are as follows:

1) It was pointed out above that the input-oriented CRS and output-oriented CRS models give the same efficiency scores; the results in the table verify this with an empirical example.
2) An efficiency score equal to 1 indicates (weak) efficiency, and if in addition all slack variables are zero, the DMU is strongly efficient. In the results of this example, 6 DMUs have an efficiency score of 1, and for these 6 DMUs all input and output slack variables are zero, which shows that all 6 DMUs are strongly efficient.
3) In an input-oriented radial model, the slack variable of at least one input is zero; in an output-oriented radial model, the slack variable of at least one output is zero. Taking the input-oriented model as an example, suppose that in the results for some DMU none of the input slack variables is zero: this would mean that room for proportionate improvement still exists and that the solution obtained is not optimal, so the case in which all input slack variables are non-zero cannot arise. In the results of this example, no DMU in the input-oriented model has non-zero slack variables for both the number of beds and the number of health technical personnel, and no DMU in the output-oriented model has non-zero slack variables for both the number of outpatient visits and the number of admissions.

2.11.2.2 Benchmarks and Projected Values (Target Values)[ii]

Table 2-8 lists the benchmarks and projected values of some of the DMUs. In the analysis results, the benchmark is denoted by Benchmark (Lambda), where Lambda is the coefficient of the DMU being referenced; when there are several reference DMUs, they are separated by semicolons. The benchmark of Anhui, for example, is "Fujian(0.6362); Xinjiang(0.7791)", which means that the projection point of Anhui on the frontier is formed as a linear combination of the inputs and outputs of Fujian and Xinjiang, with weight coefficients of 0.6362 and 0.7791 respectively.

[i] To reduce the width of the table, the decimal parts of the slack results are not retained. The same treatment is applied in the next table. To save space, the examples in this book list only part of the results.

[ii] For convenience of exposition, in this book a DMU with $\lambda*>0$ in the optimal solution is called a benchmark of the DMU under evaluation. The linear combination of the benchmarks constitutes the projected value (target value) of the DMU under evaluation. A DMU in the reference set that does not serve as a benchmark has $\lambda*=0$.

Table 2-8 Benchmarks and Projected Values of the Input-Oriented CRS and Output-Oriented CRS Models

	Input-oriented CRS					Output-oriented CRS				
DMU	Benchmark	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions	Benchmark	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions
Anhui	Fujian(0.6362); Xinjiang(0.7791)	133136	125338	74967434	4391516	Fujian(0.6737); Xinjiang(0.8251)	140997	132739	79394136	4650828
Beijing	Shanghai(0.9179)	80363	87385	104340626	2095434	Shanghai(1.0005)	87596	95250	113731343	2284025
Fujian	Fujian(1.0000)	89947	93898	72059496	3087618	Fujian(1.0000)	89947	93898	72059496	3087618

The projected value of an inefficient DMU on the frontier represents its target value for improvement. Suppose that its input and output measures reach the projected values: if the DMU were then evaluated again in the current data set (with the inputs and outputs of the other DMUs unchanged), the evaluation result would become an efficiency score equal to 1 (weak or strong efficiency).

In the MaxDEA analysis results, improvements in inputs are reported as negative numbers and improvements in outputs as positive numbers. The projected values of the inputs and outputs of the DMU under evaluation can therefore be calculated in a unified way as:

Projected value = original value + improvement value[i].

The improvement value consists of two parts: the proportionate improvement value (Proportionate Movement)[ii] and the slack improvement value (Slack Movement)[iii]. That is,

Projected value = original value + proportionate improvement value + slack improvement value.

In the input-oriented radial model, the proportionate improvement of inputs = -original value × (1 - θ*), and the proportionate improvement of outputs is 0; in the output-oriented radial model, the proportionate improvement of inputs is 0, and the proportionate improvement of outputs = original value × (φ* - 1).

[i] In some of the literature the gap between the original value and the target value of an input is called "input redundancy". "Redundancy" is a misleading name. Taken literally, redundancy simply means something superfluous: input redundancy is the superfluous part of an input, which makes no difference to the result whether it is there or not. In the projection analysis of DEA, however, the gap between the original value and the target value of an input represents the gap between the DMU under evaluation and the frontier, that is, the input factors consumed in excess of the frontier because the technical efficiency of the DMU under evaluation is low. These excess input factors are not superfluous as far as the DMU under evaluation is concerned. If the DMU were simply to cut these inputs, its output would decrease accordingly. The way to reduce or eliminate the gap to the frontier is to improve technology and management so as to raise technical efficiency.

[ii] In radial models the proportional improvement is generally called radial movement, denoting the amount by which each input (or output) is improved in the same proportion. In directional distance function models, however, this part of the improvement is not a proportional improvement but follows the direction of the direction vector. To unify the terminology for this component of improvement, it is referred to in the analysis results of MaxDEA simply as the proportional improvement.

[iii] If the projected values in the results are to be used in subsequent analysis, it is recommended that the two-stage method be selected under "Slack Calculation" on the "Options" screen. Otherwise the weakly efficient projected values obtained may still be weakly efficient, although the probability is very low.

Table 2-9 uses Anhui as an example to demonstrate the calculation of projected values by means of the improvement values.

Table 2-9 Example of Calculating Input and Output Projected Values: The Improvement-Value Method (Anhui)

	Input-oriented CRS				Output-oriented CRS			
	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions
Original value	140997	132739	63344221	4391516	140997	132739	63344221	4391516
Proportionate improvement value	-7861	-7401	0	0	0	0	3740371	259312
Slack improvement value	0	0	11623213	0	0	0	12309544	0
Projected value	133136	125338	74967434	4391516	140997	132739	79394136	4650828

In addition, the projected values of the inputs and outputs of the DMU under evaluation can also be calculated from the benchmarks and their weight coefficients:

$$\hat{x}_k = \sum_{j=1}^{n} \lambda_j^* x_j \,,\ \hat{y}_k = \sum_{j=1}^{n} \lambda_j^* y_j \,.$$

The following table likewise takes Anhui as an example to demonstrate the calculation of strongly efficient projected values by means of the benchmark values. Because the largest original value has 8 digits, the weight coefficients must be retained to 9 decimal places in order to avoid calculation error. The projected values calculated by the two methods are the same.

Table 2-10 Example of Calculating Input and Output Projected Values: The Benchmark-Value Method (Anhui)

	Input-oriented CRS				Output-oriented CRS			
	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions	Number of beds	Health technical personnel	Outpatient visits	Inpatient admissions
Fujian original value	89947	93898	72059496	3087618	89947	93898	72059496	3087618
Fujian weight	0.636154651				0.673718522			
Xinjiang original value	97436	84202	37383273	3115413	97436	84202	37383273	3115413
Xinjiang weight	0.779130551				0.825136911			
Projected value	133136	125338	74967434	4391516	140997	132739	79394136	4650828

2.11.2.3 Weight Coefficients of the Inputs and Outputs

Select the multiplier model (menu Step 2: Run Model-Run Multiplier Model). In the "Weights (Dual Values)" table of the results window that displays the analysis results, the weights of the inputs and outputs are denoted by Weight (measure name); Weight (number of beds), for example, denotes the weight coefficient of the input measure "number of beds".

2.11.2.4 Relationship between the Dual Solutions of the Envelopment Model and the Multiplier Model

The solution of the envelopment model corresponds to the dual solution of the multiplier model, and the solution of the multiplier model corresponds to the dual solution of the envelopment model. In the dual solution of the envelopment model, the dual prices of the inputs and outputs correspond to the input and output weight coefficients in the multiplier model. Since the multiplier model often has multiple optimal solutions, the two may not be numerically identical.

2.11.2.5 Scale Efficiency

In the "Returns to Scale" option, if "Efficiency or TFP index decomposition (including scale efficiency / scale effect)" is selected, the software runs the CRS model and then the VRS model, and reports the efficiency scores of the two models together with the scale efficiency (Scale Efficiency) or scale effect (Scale Effect) calculated from those two efficiency scores[i]. The scale efficiency obtained from the envelopment model and from the multiplier model is the same.

In the results window of MaxDEA, the three efficiency scores displayed are:

1) Technical Efficiency Score(CRS): the technical efficiency score obtained from the CRS model. To distinguish it from the technical efficiency score obtained from the VRS model, the technical efficiency (Technical Efficiency, TE) obtained from the CRS model is usually called overall technical efficiency.
2) Pure Technical Efficiency Score(VRS): the technical efficiency score obtained from the VRS model, usually called pure technical efficiency (Pure Technical Efficiency Score, PTE).
3) Scale Efficiency Score/ Scale Effect Score: the scale efficiency (Scale Efficiency, SE) or the scale effect (Scale Effect, SE).

Table 2-11 lists the overall technical efficiency (TE), pure technical efficiency (PTE) and scale efficiency (SE) of some provinces. The relationship among the three is:

Scale efficiency (SE) = overall technical efficiency (TE) / pure technical efficiency (PTE)

Table 2-11 Scale Efficiency

	Input-oriented model			Output-oriented model		
DMU	**Overall technical efficiency (TE)**	**Pure technical efficiency (PTE)**	**Scale efficiency (SE)**	**Overall technical efficiency (TE)**	**Pure technical efficiency (PTE)**	**Scale efficiency (SE)**
Anhui	0.9442	0.9549	0.9889	0.9442	0.9568	0.9869
Beijing	0.9174	0.9185	0.9989	0.9174	0.9175	0.9999

[i]According to the definition of scale efficiency in the literature, the radial distance should be used to calculate "scale efficiency" (Scale Efficiency). When other distance types are used, the resulting measure is called the "scale effect" (Scale Effect) in MaxDEA.

Fujian	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Gansu	0.8324	0.8538	0.9750	0.8324	0.8473	0.9825
Guangdong	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Guangxi	0.9901	0.9961	0.9940	0.9901	0.9964	0.9937
Guizhou	0.9631	0.9708	0.9920	0.9631	0.9698	0.9931
Hainan	0.8331	0.9415	0.8849	0.8331	0.9323	0.8936
Hebei	0.9039	0.9391	0.9625	0.9039	0.9452	0.9563
Henan	0.8826	0.9615	0.9180	0.8826	0.9698	0.9101
Heilongjiang	0.6893	0.6900	0.9990	0.6893	0.7020	0.9820
Hubei	0.9469	0.9704	0.9759	0.9469	0.9724	0.9738

2.12 Common Misunderstandings about Projected Values

In the results of a DEA model, the projected value (target value) is the projection of the DMU under evaluation onto the frontier, and the relative gap between the current value of a DMU and its projected value represents the degree of inefficiency of each of its inputs and outputs.

The projected value is also called the target value, and a misunderstanding that readily arises is that as long as the input and output values of an inefficient DMU are changed to the projected values, that DMU will become an efficient DMU. This is a mechanical reading of the results. From the computational procedure alone, if the input and output values of an inefficient DMU are modified to its projected values (assuming that the other DMUs remain unchanged) and the efficiency score is recalculated, the DMU will indeed become efficient. In actual production processes, however, if the quantity of an input used by a DMU is simply reduced, the quantity of its output must fall accordingly; and if the quantity of the output of a DMU is to be increased on its own, its inputs must be increased accordingly. To require a DMU, purely on the basis of the projection results, to reduce its inputs while holding outputs constant (or even increasing them), or to increase its outputs while holding inputs constant (or even reducing them), is impossible to achieve.

The gap between the current value and the projected value is the result of a DMU's inefficiency, not the cause of it. DEA analysis itself cannot identify the causes of inefficiency, and therefore cannot directly indicate the way to improve efficiency. A DMU may be inefficient because its production technology is backward, because its management is poor, or for other reasons. DEA analysis reports the state of a DMU's inefficiency (the efficiency score and the gap to the frontier); it does not report the causes of inefficiency, still less can it directly indicate the way to improve efficiency.

The results of DEA analysis can nonetheless give indirect guidance for an inefficient DMU to improve its efficiency. The analysis results give, for example, the benchmark of the inefficient DMU. The inefficient DMU can analyse its benchmark DMU in depth, identify the technological and managerial gaps between itself and the benchmark, and determine its own way of improving efficiency. When an inefficient DMU improves its technology and management and its technical efficiency rises, the gap between that DMU and the frontier will narrow or disappear.

In short, the gap between an inefficient DMU and the frontier is the result of its technical inefficiency,

not the cause of it. Reductions in inputs or increases in outputs are the expected outcome after an inefficient DMU has achieved technical efficiency by improving its technology and management. The way to raise technical efficiency is to improve technology and management, not to reduce inputs or increase outputs.

2.13 Relationship between the Output Results of the Envelopment Model and the Multiplier Model

2.13.1 Graphical Illustration of the Projections of the Envelopment Model and the Multiplier Model

Besides the efficiency score, the output results of the envelopment model include the projection point of the DMU under evaluation on the frontier (the benchmark and the projected values). For instance, in Figure 2-11 the point E has a projection point E' on the frontier, which the envelopment model results express as E' = 0.5 A + 0.5B. The projection point of the DMU under evaluation on the frontier is expressed as a linear combination of the vertices (A and B) of that segment AB of the frontier. Any point between A and B can be expressed as P = λ1A + λ2B, with λ1 + λ2 = 1, λ1 ≥ 0 and λ2 ≥ 0.

Besides the efficiency score, the output results of the multiplier model include the weights of the inputs and outputs and the free variable (u* or v*). The multiplier model results for point E are a weight of 0.2 for input x, a weight of 0.1 for output y and a free variable of -0.4. The results given by the multiplier model correspond to the equation of the hyperplane (that is, the line AB) on which the projection point E' of point E on the frontier lies, namely

-2x + y + 4 = 0. Multiplying both sides of the equation by 0.1 gives -0.2x + 0.1y + 0.4 = 0.

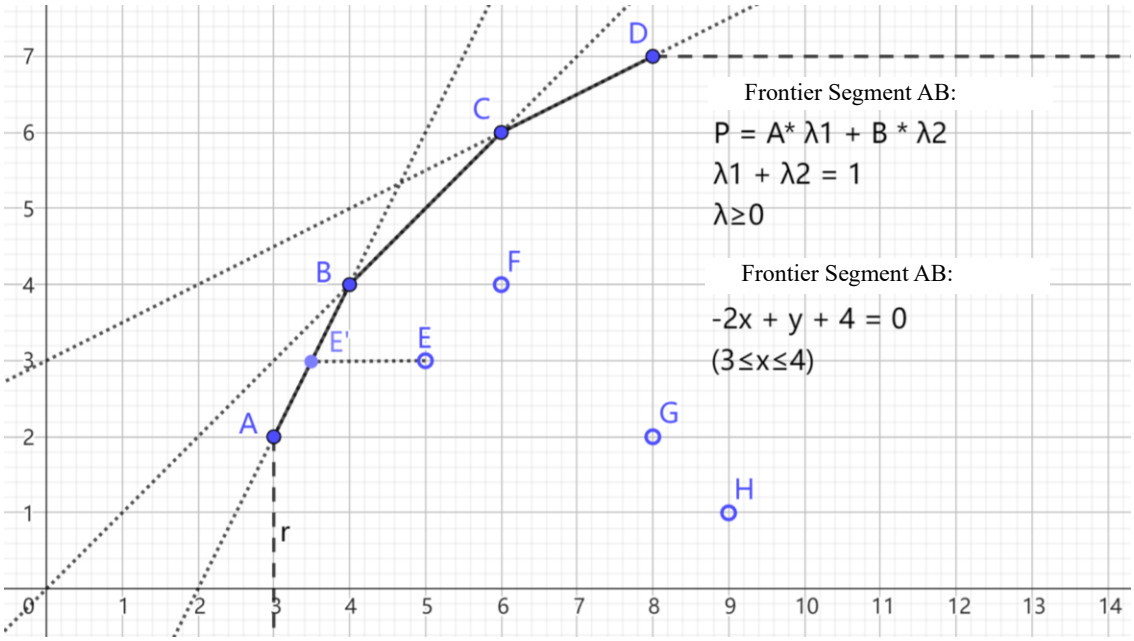


Figure 2-11 Relationship between the Output Results of the Envelopment Model and the Multiplier Model (Single-Input, Single-Output Input-Oriented VRS Model)

2.13.2 Why the Multiplier Model Has Multiple Optimal Solutions

For instance, in Figure 2-11 the point F has the projection point B, and B lies on both of the hyperplanes given by the lines AB and BC, so F has two optimal solutions. The projection point of G is A, which likewise gives two optimal solutions. Moreover, the vertices on the frontier (A, B, C and D) each lie on two hyperplanes at the same time, so each of them has multiple optimal solutions.

The larger the number of input and output measures, the more often multiple optimal solutions occur and the larger the number of optimal solutions. In a VRS multiplier model with three inputs and three outputs, for example, the number of optimal solutions may be as many as several hundred (a single vertex on the frontier may lie on several hundred hyperplanes at the same time).

2.13.3 Why a Weight of Zero Appears in the Multiplier Model Results

For instance, in Figure 2-11 the projection point of the point H lies on r, the perpendicular from point A to the X axis, and the equation of that perpendicular is $-x + 0y + 3 = 0$. The coefficient of y is 0, so the corresponding output weight is 0. When the frontier on which the projection point lies is parallel to a coordinate axis, the weight of the corresponding input or output is 0.

Chapter 3 Returns to Scale

3.1 Types of Returns to Scale in DEA Models

In DEA models, the specification of returns to scale (RTS) determines the shape of the frontier. In the single-input, single-output case, the frontiers of the CRS and VRS models introduced in the previous chapter are respectively a ray and a convex curve. As shown in the figure, in the CRS model the frontier is the ray OB, and point B is the only efficient DMU (Figure 3-1); whereas in the VRS model the frontier is the convex curve MABD, and there are three efficient DMUs, A, B and D (Figure 3-2).

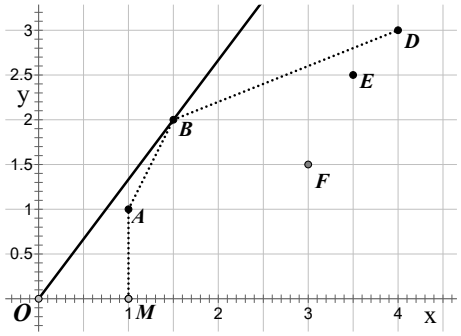


Figure 3-1 CRS Model Frontier

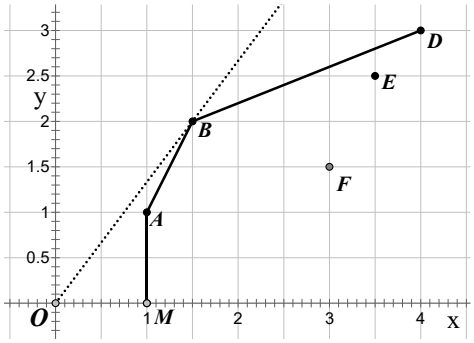


Figure 3-2 VRS Model Frontier (Input-Oriented)

How, then, should the RTS type be chosen when a DEA model is built? If the production technology under evaluation is known to be of the CRS type, or if it is of the VRS type but the production scale of the sample is known to lie in the CRS stage, the CRS model should be chosen. In actual production processes, however, such situations often do not occur: the returns to scale of the production technology are not CRS, or it cannot be determined whether the production scale of the sample lies in the CRS stage. In this case, if the CRS model is adopted, the technical efficiency obtained is not entirely pure technical efficiency but an overall technical efficiency that contains a scale efficiency component.

Generally, the returns to scale of a production technology pass through three stages in succession: increasing returns to scale (IRS), constant returns to scale (CRS) and decreasing returns to scale (DRS). If it cannot be determined which stage the sample lies in, or if the sample is known to span the various RTS stages, the VRS model should be chosen when technical efficiency is evaluated. The technical efficiency obtained from the VRS model is pure technical efficiency.

If the sample is known to lie in the IRS or CRS stage (and certainly not in the DRS stage;

non-decreasing returns to scale, NDRS), or is known to lie in the CRS or DRS stage (and certainly not in the IRS stage; non-increasing returns to scale, NIRS)[i], what shape does the frontier take in each case?

The frontier MABD of the VRS model can be divided into three parts: MAB (excluding point B) is the IRS stage, point B is the CRS stage, and BD (excluding point B) is the DRS stage. The frontier of the NDRS model contains no DRS stage, so the NDRS frontier consists of two parts: MAB (excluding point B) is the IRS stage, and the ray r with point B as its endpoint is the CRS stage (Figure 3-3). The frontier of the NIRS model contains no IRS stage and consists of two parts: OB is the CRS stage, and BD (excluding point B) is the DRS stage (Figure 3-4).

The VRS model adds the constraint $\Sigma\lambda=1$ to the CRS model. Any point on the VRS frontier curve ABD[ii] can be formed as a linear combination of the efficient DMUs (A, B and D) under the constraint that the coefficients sum to 1.

In Figure 3-3 the curve MAB and the ray r constitute the frontier of the NDRS model, where on segment AB (IRS stage) the sum of the linear combination coefficients equals 1, and on the ray segment r (CRS stage) the sum of the linear combination coefficients is greater than or equal to 1 (only at point B does it equal 1); the linear combination coefficient constraint of the NDRS model is therefore greater than or equal to 1, that is, $\Sigma\lambda\geq1$[iii].

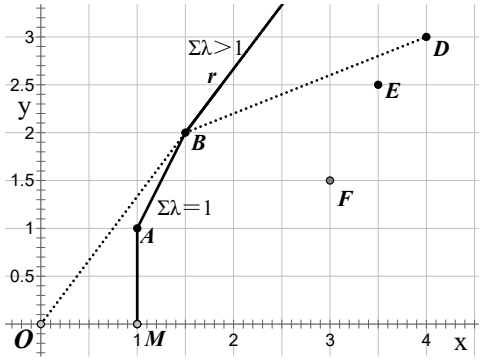


Figure 3-3 NDRS Model Frontier (Input-Oriented)

In Figure 3-4 the curve OBD constitutes the frontier of the NIRS model, where on segment OB (CRS stage) the sum of the linear combination coefficients is less than or equal to 1 (only at point B does it equal 1), and on segment BD (DRS stage) the sum of the linear combination coefficients = 1; the linear combination coefficient constraint of the NIRS model is therefore less than or equal to 1, that is, $\Sigma\lambda\leq1$.

[i]In some of the literature the NDRS model is called the IRS model and the NIRS model the DRS model, which is an inaccurate naming, because the NDRS model contains both the IRS and the CRS ranges and the NIRS model contains both the CRS and the DRS ranges.

[ii]The segment MA parallel to the coordinate axis is the weakly efficient frontier; its actual strongly efficient projection point is point A, and the sum of the linear combination coefficients also equals 1.

[iii]Any point on the ray OB can be constructed by multiplying point B by a coefficient, written S=kB. Below point B the coefficient k is less than 1; beyond point B the coefficient k is greater than 1.

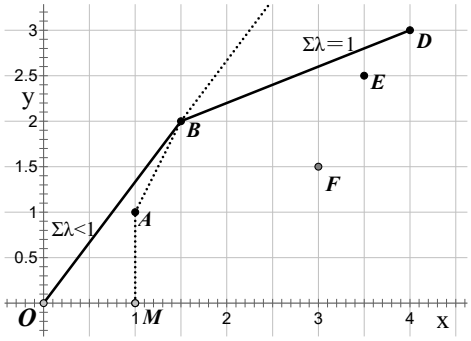


Figure 3-4 NIRS Model Frontier (Input-Oriented)

By the duality theory of linear programming, the multiplier models for the various RTS types can be derived (Table 3-1).

Table 3-1 DEA Models of Different RTS Types

	Envelopment model		**Multiplier model**	
	Input-oriented	**Output-oriented**	**Input-oriented**	**Output-oriented**
CRS	$\min\ \theta$ $s.t.\ \theta x_0 - X\lambda \geq 0$ $Y\lambda \geq y_0$ $\lambda, s^-, s^+ \geq 0$	$\max\ \phi$ $s.t.\ X\lambda \leq x_0$ $-Y\lambda + \phi y_0 \leq 0$ $\lambda, s^-, s^+ \geq 0$	$\max \mu' y_0 - \mu_0$ $s.t.\ \nu' x_0 = 1$ $-\nu' X + (\mu' Y - \mu_0) \leq 0$ $\mu, \nu \geq 0\ (or\ \mu, \nu \geq \varepsilon)$ $\mu_0 = 0$	$\min \nu' x_0 + \nu_0$ $s.t.\ \mu' y_0 = 1$ $(\nu' X + \nu_0) - \mu' Y \geq 0$ $\mu, \nu \geq 0\ (or\ \mu, \nu \geq \varepsilon)$ $\nu_0 = 0$
VRS	$e\lambda = 1$	$e\lambda = 1$	μ_0 *free*	ν_0 *free*
NDRS	$e\lambda \geq 1$	$e\lambda \geq 1$	$\mu_0 \leq 0$	$\nu_0 \leq 0$
NIRS	$e\lambda \leq 1$	$e\lambda \leq 1$	$\mu_0 \geq 0$	$\nu_0 \geq 0$

e denotes a row vector all of whose elements equal 1, that is, e=(1,1,…,1). In the formulations of different studies, the free variables of the multiplier model (μ0 and ν0) are preceded sometimes by a positive sign and sometimes by a negative sign, but the signs in the constraints change accordingly, so the results are the same.

If it is already clear which DMUs in the sample are in the IRS stage, which in the CRS stage and which in the DRS stage, the NDRS, CRS and NIRS models can be built separately for the respective groups. If only the production technology is known to be VRS while the RTS status of the individual DMUs in the sample is unclear, the VRS model is chosen to calculate pure technical efficiency.

References for this section: (Rolf Färe, Grosskopf, & Lovell, 1983; Grosskopf, 1986; William W Cooper et al., 2007; J. Zhu, 2009).

3.2 Determining the RTS Status of a DMU

3.2.1 Determining RTS Status through the Envelopment Model

The preceding section discussed how to choose a DEA model according to the returns-to-scale properties of the production technology or of the sample. This section discusses how, given that the production technology under study and the sample exhibit variable returns to scale, the RTS stage of each DMU can be determined, that is, whether it is in the IRS, CRS or DRS stage.

As shown in the preceding section, DEA models for production technologies with different returns to scale can be constructed by imposing different constraints on the "sum of the linear combination coefficients". When the production technology exhibits variable returns to scale and the RTS stage of the sample is unknown, the VRS model should be built. In the results of the VRS model, however, the "sum of the linear combination coefficients" always equals 1, so it is clear that this sum cannot be used to determine which RTS stage a DMU is in.

It was noted above that the VRS model should be built when a production technology with variable returns to scale is studied and the RTS status of the sample is unknown. Suppose that a CRS model is also built and that the frontiers of the CRS and VRS models are placed in the same coordinate system; the result is shown in Figure 3-5 below, where the frontier of the CRS model (ray OC) and the frontier of the VRS model (the solid portion) are placed together.

Consider the frontier curve of the VRS model: point A and the projection of F lie in the IRS stage, points B and C and the projection of G lie in the CRS stage, and point D and the projection of E lie in the DRS stage. The projections of points A and F onto the CRS frontier curve (ray OC) lie below point B, and their "sum of the linear combination coefficients" in the CRS model is less than 1; the projections of points B, C and G onto the CRS frontier curve lie on segment BC, and their "sum of the linear combination coefficients" in the CRS model equals 1; the projections of points D and E onto the CRS frontier curve lie above point C, and their "sum of the linear combination coefficients" in the CRS model is greater than 1.

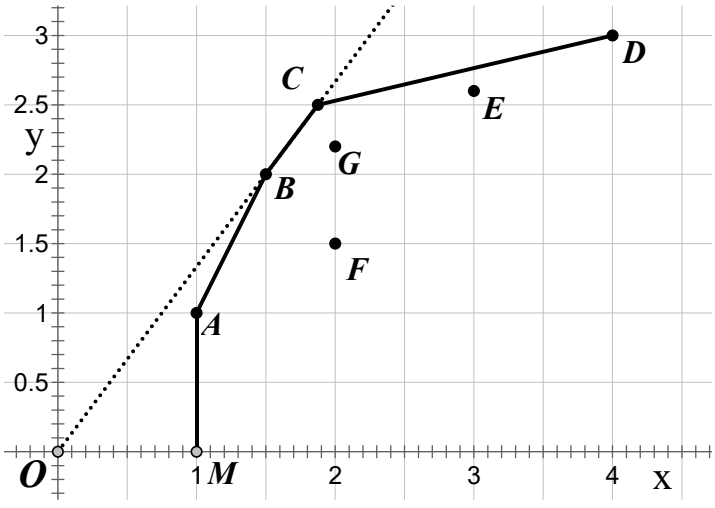


Figure 3-5 Determining the RTS Status of a DMU

Summarising these patterns, it seems that the CRS model can readily be borrowed to determine the

RTS status of each DMU in the VRS model: if, in the results for the DMU under evaluation,

1) Σλ*<1, the DMU is in the IRS stage;

2) Σλ*=1, the DMU is in the CRS stage;

3) Σλ*>1, the DMU is in the DRS stage.

However, the frontier curve of the CRS model (ray OC) contains two observed points, namely points B and C. Any point on ray OC can be constructed linearly from point B alone, from point C alone, or from the two points jointly, that is,

S=λB or S=λC or S=λ1B+λ2C

This means that multiple optimal solutions necessarily exist in the CRS model. In Figure 3-5 the projection of G onto the CRS frontier lies between B and C: if this projection is constructed from point B, then λ>1; if it is constructed from point C, then λ<1; and if it is constructed jointly from points B and C, then λ<1, λ=1 and λ>1 may all occur. In fact, however, G is in the CRS stage, which shows that the above method of determining the RTS status of a DMU is not entirely correct. To avoid such misjudgements, the multiple-solution problem must be taken into account, and the method is amended as follows:

RTS determination method 1:

1) if Σλ*<1 in all optimal solutions of the CRS model, the DMU is in the IRS stage;

2) if Σλ*>1 in all optimal solutions of the CRS model, the DMU is in the DRS stage;

3) if Σλ*=1 in any one of the optimal solutions of the CRS model, the DMU is in the CRS stage.

Solving for all optimal solutions of a DEA model is not an easy task. Establishing the above conditions does not require all optimal solutions. In actual computation, it suffices, after the CRS model has been solved, to solve two additional models so as to obtain the maximum and the minimum value of Σλ respectively.

Taking the input-oriented model as an example, the linear programming formulation for the maximum (minimum) value of Σλ is

$$\begin{aligned}
&\max\ (\min)\ \sum_{j=1}^{n}\lambda_j \\
&s.t.\ \sum_{j=1}^{n}\lambda_j\theta_j^* x_{ij} \le \theta_k^* x_{ik} \\
&\sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{rk} \\
&\lambda \ge 0 \\
&i=1,2\ldots,m;\ r=1,2\ldots,q;\ j=1,2\ldots,n \quad (3.1)
\end{aligned}$$

In the above formulation, if Max(Σλ) <1, then Σλ is less than 1 in all optimal solutions; if Min(Σλ) >1, then Σλ is greater than 1 in all optimal solutions.

RTS determination method 2:

1) if Max (Σλ) <1, the DMU is in the IRS stage;

2) if Min (Σλ) >1, the DMU is in the DRS stage;

3) if neither of the above two cases applies, the DMU is in the CRS stage.

Compared with obtaining all optimal solutions, the above method of solving for the maximum and minimum values of Σλ is clearly more convenient. This method, however, still requires two additional linear programming formulations, which increases the computational burden of solving the programmes.

Judging RTS status from Σλ in a single optimal solution may misclassify a DMU that is actually in the CRS stage as being in the IRS or DRS stage. From Figure 3-5 it can be observed that the frontier of the CRS model and the frontier of the VRS model overlap in a region, namely the CRS portion of the VRS frontier curve (segment BC). If an efficient DMU is in the CRS stage, it must lie within the CRS region of the frontier curve. Combining this with the method of calculating scale efficiency introduced in the previous chapter, it follows that if a DMU is in the CRS stage, its scale efficiency = 1; conversely, if the scale efficiency of a DMU = 1, the DMU is in the CRS stage. From this rule it follows that

RTS determination method 3:

1) if scale efficiency = 1, the DMU is in the CRS stage;

2) if scale efficiency <1 and Σλ*<1 in any optimal solution, the DMU is in the IRS stage;

3) if scale efficiency <1 and Σλ*>1 in any optimal solution, the DMU is in the DRS stage.

3.2.2 Determining RTS Status through the Multiplier Model

By the duality theory of linear programming, RTS status can also be determined through the multiplier model.

Reasoning 1:

Taking the input-oriented model as an example, suppose that the projected value of DMU K on the weak efficient frontier is ($\theta_k^* x_{ik}, y_{rk}$), then $\sum_{r=1}^{q} \mu_r y_{rk} - \sum_{i=1}^{m} \nu_i \theta_k^* x_{ik} - \mu_{0k} = 0$ is a supporting hyperplane of the frontier passing through the projected point ($\theta_k^* x_{ik}, y_{rk}$).

1) If the projected point is in a constant-returns-to-scale (CRS) state, then when the inputs and outputs of the projected point are increased simultaneously and proportionally by a very small amount (the proportion of increase being a sufficiently small positive number ε)[i], the projected point after displacement will still lie on a supporting hyperplane passing through that point, that is, it satisfies

$$\sum_{r=1}^{q} \mu_r (1+\varepsilon) y_{rk} - \sum_{i=1}^{m} \nu_i \theta_k^* (1+\varepsilon) x_{ik} - \mu_{0k}$$

$$= (1+\varepsilon) \sum_{r=1}^{q} \mu_r y_{rk} - (1+\varepsilon) \sum_{i=1}^{m} \nu_i \theta_k^* x_{ik} - \mu_{0k} = 0.$$

[i]A small proportional increase in both the inputs and the outputs of the projected point is equivalent to extending the projected point slightly along the line joining the origin to that point. If the projected point is in a constant-returns-to-scale (CRS) state, the extended point still lies on one of the supporting planes of the projected point (the supporting plane passing through the origin); if the projected point is in an increasing-returns-to-scale (IRS) state, the extended point lies to the right of all the supporting planes of the projected point; and if the projected point is in a decreasing-returns-to-scale (DRS) state, the extended point lies to the left of all the supporting planes of the projected point.

From $\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}=0$ it follows that

$(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-(1+\varepsilon)\mu_{0k}=0$;

therefore, $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}=0$ holds if and only if $\varepsilon\mu_{0k}=0$, that is, $\mu_{0k}=0$.

Where multiple optimal solutions exist, it suffices that any one optimal solution satisfies $\mu_{0k}^*=0$ for the projected point to be in the CRS stage of the VRS frontier, with an RTS status of CRS.

2) If the projected point is in an increasing-returns-to-scale (IRS) state, then after displacement it will lie to the right of the supporting hyperplane passing through that point, that is, it satisfies $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}<0$. The corresponding condition is $\varepsilon\mu_{0k}<0$, that is, $\mu_{0k}<0$. If several supporting hyperplanes pass through the projected point (that is, the linear programme of the multiplier model has multiple optimal solutions), the projected point after displacement will lie to the right of all the supporting hyperplanes, that is, in all optimal solutions $\mu_{0k}^*<0$.

3) If the projected point is in a decreasing-returns-to-scale (DRS) state, then after displacement it will lie to the left of the supporting hyperplane passing through that point, that is, it satisfies $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}>0$. If several supporting hyperplanes pass through the projected point (that is, the linear programme of the multiplier model has multiple optimal solutions), the projected point after displacement will lie to the left of all the supporting hyperplanes, that is, in all optimal solutions $\mu_{0k}^*>0$.

Similar reasoning applies to the output-oriented model.

Reasoning 2:

If marginal output is greater than average output, the DMU is in an increasing-returns-to-scale (IRS) state; if marginal output equals average output, it is in a constant-returns-to-scale (CRS) state; and if marginal output is less than average output, it is in a decreasing-returns-to-scale (DRS) state. Therefore the RTS status can be determined by comparing the marginal product (MP) and the average product (AP) of the projected point of the DMU under evaluation on the weak efficient frontier.

Now regard $\sum_{i=1}^{m}v_i\theta_k^* x_{ik}$, $\sum_{r=1}^{q}\mu_r y_{rk}$ as a virtual input and a virtual output respectively[i], that

[i]The virtual input and the virtual output can be understood as follows: the virtual input and the virtual output constructed from the weight coefficients in the multiplier model correspond to the constraints on proportional changes in inputs or outputs in the envelopment model.

is, $x=\sum_{i=1}^{m}v_i\theta_k^*x_{ik}$, $y=\sum_{r=1}^{q}\mu_r y_{rk}$.

$\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^*x_{ik}-\mu_{0k}=0$, written for simplicity as $y-x-u_{0k}=0$, is a supporting hyperplane of the frontier passing through the projected point, so the marginal output of the projected point can be calculated as

$$MP=\frac{dy}{dx}=1$$

From

$$\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^*x_{ik}-\mu_{0k}=0 \quad \text{and} \quad \sum_{i=1}^{m}v_i\theta_k^*x_{ik}=1$$

the average output at this input point is

$$AP=\frac{\sum_{r=1}^{q}\mu_r y_{rk}}{\sum_{i=1}^{m}v_i\theta_k^*x_{ik}}=\sum_{r=1}^{q}\mu_r y_{rk}=1+\mu_{ok}$$

Combining this with the rules for determining RTS status, the method for determining RTS from the input-oriented multiplier model is obtained (Table 3-2).

Table 3-2 Methods for Determining RTS Status in the Input-Oriented Multiplier Model

	Comparison of MP and AP	**μ_0**
IRS	MP>AP in all solutions	in all solutions $\mu_0^*<0$
CRS	MP=AP in any solution	in any solution $\mu_0^*=0$
DRS	MP<AP in all solutions	in all solutions $\mu_0^*>0$

Similar derivations show that in the output-oriented multiplier model MP=1, $AP=\frac{1}{1-v_{ok}}$, from which the method for determining RTS status is obtained (Table 3-3).

Table 3-3 Methods for Determining RTS Status in the Output-Oriented Multiplier Model

	Comparison of MP and AP	**v_0**
IRS	MP>AP in all solutions	in all solutions $v_0^*<0$
CRS	MP=AP in any solution	in any solution $v_0^*=0$
DRS	MP<AP in all solutions	in all solutions $v_0^*>0$

It should be noted that method 2 for the multiplier model is consistent with method 2 for the

envelopment model. In the envelopment model(3.1) the right-hand side of the formulation is the projected value of the DMU under evaluation; therefore, in the multiplier model, an inefficient DMU must first be projected[i] to obtain its projected value, after which the maximum and minimum values of μ0 or ν0 are solved. Taking the input-oriented model as an example[ii],

$$\max(\min)\ \mu_0$$

$$s.t.\ \sum_{r=1}^{q}\mu_r y_{rj} - \sum_{i=1}^{m}\nu_i\theta_j^* x_{ij} - \mu_0 \le 0$$

$$\sum_{i=1}^{m}\nu_i\theta_k^* x_{ik} = 1$$

$$\nu \ge 0; \mu \ge 0;\ \mu_0\ free$$

$$i = 1,2\ldots,m;\ r = 1,2\ldots,q;\ j = 1,2\ldots,n \quad (3.2)$$

Table 3-4 summarizes the various methods for determining RTS status in the envelopment model and the multiplier model. The rule adopted by MaxDEA for determining RTS status is method 3.

References for this section:(Banker et al., 1984; Banker & Thrall, 1992).

[i] A weakly efficient projection is used here, that is, original value + radial improvement.

[ii] θ * is the efficiency score of each DMU; the efficiency score obtained from the multiplier model equals that obtained from the envelopment model, that is, $\theta_k^* = \sum_{r=1}^{q}\mu_r^* y_{rk} - \mu_0^*$.

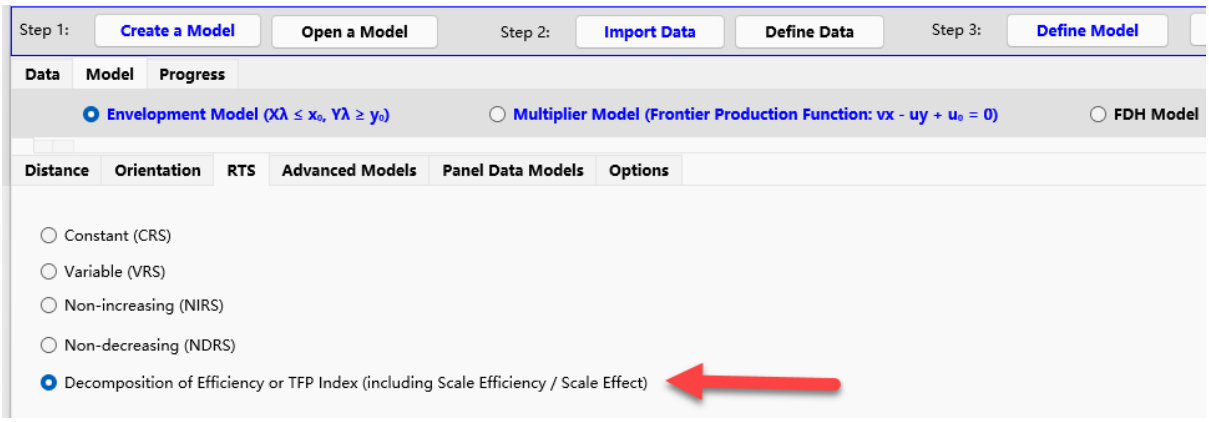
Table 3-4 Summary of Methods for Determining RTS Status

	Envelopment model		Multiplier model	
	Input-oriented	**Output-oriented**	**Input-oriented**	**Output-oriented**
Determination Model	$\min\ \theta$ $s.t.\ x_0\theta - X\lambda \geq 0$ $Y\lambda \geq y_0$ $\lambda, s^-, s^+ \geq 0$	$\max\ \phi$ $s.t.\ X\lambda \leq x_0$ $y_0\phi - Y\lambda \leq 0$ $\lambda, s^-, s^+ \geq 0$	$\max\ y_0^T\mu - \mu_0$ $s.t.\ x_0^T\nu = 1$ $-X^T\nu + (Y^T\mu - \mu_0) \leq 0$ $\mu, \nu \geq 0$ $\mu_0\ free$	$\min\ x_0^T\nu + \nu_0$ $s.t.\ y_0^T\mu = 1$ $(X^T\nu + \nu_0) - Y^T\mu \geq 0$ $\mu, \nu \geq 0$ $\nu_0\ free$
Method 1	1) eλ*<1 in all optimal solutions: IRS; 2) eλ*>1 in all optimal solutions: DRS; 3) eλ*=1 in any optimal solution: CRS.		1) μ0*<0 (ν0*<0) in all optimal solutions: IRS; 2) μ0*>0 (ν0*>0) in all optimal solutions: DRS; 3) μ0*=0 (ν0*=0) in any optimal solution: CRS.	
Method 2	1) eλ*max <1: IRS; 2) eλ*min>1: DRS; 3) otherwise: CRS.		1) μ0*,max < 0 (ν0*,max < 0): IRS; 2) μ0*,min > 0 (ν0*,min > 0): DRS; 3) otherwise: CRS.	
Method 3	1) scale efficiency = 1: CRS; 2) scale efficiency <1 and eλ*<1 in any optimal solution: IRS; 3) scale efficiency <1 and eλ*>1 in any optimal solution: DRS.		1) scale efficiency = 1: CRS; 2) scale efficiency <1 and μ0*<0 (ν0*<0) in any optimal solution: IRS; 3) scale efficiency <1 and μ0*>0 (ν0*>0) in any optimal solution: DRS.	

X is an m×n input matrix, Y is a q×n output matrix, λ is an n×1 column vector (benchmark coefficients), ν is an m×1 column vector (input weights) and μ is a q×1 column vector (output weights)

3.3 Graphical Illustration of Solving for Scale Efficiency and RTS Status in MaxDEA

Under "Returns to Scale", select "Decomposition of Efficiency or TFP Index (including Scale Efficiency / Scale Effect)".



3.4 The Relationship between RTS Status and Model Orientation

The RTS status of a DMU reported in the results of a DEA analysis is in essence the RTS status of its frontier projection. Different model orientations yield different frontier projections and hence different RTS statuses. As shown in the figure below, point F is found to exhibit increasing returns to scale under the input-oriented model but decreasing returns to scale under the output-oriented model.

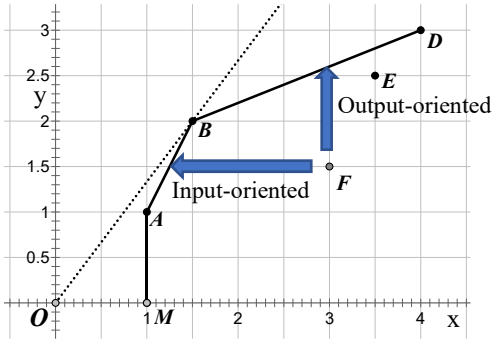


Figure 3-6 Relationship of RTS Status and Scale Elasticity to Model Orientation

Chapter 4 Distance Functions in DEA

In the DEA models introduced above, an inefficient DMU is improved by contracting (expanding) all inputs (outputs) proportionally, and such models are called radial DEA models. The radial distance function is one of the distance function types used in DEA models; besides it, the commonly used distance function types also include the farthest distance function to the frontier (SBM model), the minimum distance function to the strong efficient frontier, the directional distance function, the minimum distance function to the weak efficient frontier, the hybrid distance function and the cost function.

4.1 Radial Distance

Radial distance functions have already been explained in detail in Chapters 2 and 3 of this book and are not repeated here.

Two terms are additionally explained here: the Shepherd input distance (Shepherd Input Distance) and the Shepherd output distance (Shepherd Output Distance), together with their relationship to Farrell technical efficiency(Farrell, 1957).

In the input-oriented radial model(2.10) the Shepherd input distance is expressed as 1/θ, and its value range is 1 to ∞. Farrell technical efficiency is θ.

In the output-oriented radial model(2.14) the Shepherd output distance is expressed as 1/φ, and its value range is 0 to 1. Farrell technical efficiency is also 1/φ.

It should be noted that in some of the literature the analysis results report input or output distance values rather than Farrell technical efficiency values. In input-oriented models the two are reciprocals of each other; in output-oriented models they are equal.

When undesirable outputs are included, the radial models under the various orientations are as follows (taking VRS as an example).[i]

1) Input-oriented

$$\begin{aligned} &\max \theta \\ &s.t.\ X\lambda - \theta x_k \le 0 \\ &\quad Y\lambda \ge y_k \\ &\quad B\lambda \le b_k \\ &\quad e\lambda = 1 \\ &\quad \lambda \ge 0 \end{aligned} \quad (4.1)$$

2) Output-oriented

$$\begin{aligned} &\max \phi = 1+\beta \\ &s.t.\ X\lambda \le x_k \\ &\quad Y\lambda - \beta y_k \ge y_k \\ &\quad B\lambda + \beta b_k \le b_k \\ &\quad e\lambda = 1 \\ &\quad \lambda \ge 0 \end{aligned} \quad (4.2)$$

[i]To make the formulas more concise, matrix notation is used.

3) Non-oriented

$$\max \frac{1-\beta}{1+\beta}$$

$$s.t.\ X\lambda+\beta x_k \le x_k$$
$$Y\lambda-\beta y_k \ge y_k$$
$$B\lambda+\beta b_k \le b_k$$
$$e\lambda=1$$
$$\lambda \ge 0 \quad (4.3)$$

The improvement proportions of the inputs, the desirable outputs and the undesirable outputs are equal.

4.2 Farthest Distance to the Frontier (SBM Model)

4.2.1 The SBM Model

In radial DEA models, the measurement of the degree of inefficiency covers only the proportion by which all inputs (outputs) are contracted (expanded) proportionally. For an inefficient DMU, the gap between its current state and its projected value comprises, besides the proportional improvement, the slack improvement; yet the slack improvement is not reflected in the measurement of the efficiency score. For this reason,Kaoru Tone (2001) proposed the SBM model (slack based measure, SBM).

$$\min \ \rho = \frac{1-\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.4)$$

In fact, even earlier,J. T. Pastor, Ruiz, & Sirvent (1999) had proposed exactly the same model and named it the ERM model (Enhanced Russell Graph Efficiency Measure). In most of the literature that followed, however, the name SBM is used[i].

The SBM(ERM) model uses ρ* to denote the efficiency score of the evaluated DMU; it measures inefficiency from the input and the output side simultaneously and is therefore called a non-oriented model. In the non-oriented SBM(ERM) model, neither the input data nor the output data may contain zeros.

In the SBM(ERM) model, input and output inefficiency are respectively reflected as

$$\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik} \ , \ \frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}$$

If the efficiency score (ρ*) of the SBM(ERM) model equals 1, the evaluated DMU is strongly efficient, and the weak efficiency problem of radial models does not arise. The projected value (target value) of the evaluated DMU K is

$$\hat{x}_k = x_k - s^-; \ \hat{y}_k = y_k + s^+$$

In radial models, inefficiency is measured by the degree to which all inputs (outputs) can be contracted (expanded) proportionally, whereas in the SBM(ERM) model inefficiency is measured by the average proportion by which the individual inputs (outputs) can be contracted (expanded).

The advantage of the SBM(ERM) model is that it solves the problem that the measurement of inefficiency in radial models does not include slack variables, but the SBM(ERM) model also has evident drawbacks. The objective function of the SBM(ERM) model minimizes the efficiency score ρ, that is, it

[i]In this model the variables s- and s+ both appear in the objective function and are not slack variables, so it is not reasonable to name this model SBM (slacks based measure).

maximizes input and output inefficiency. Viewed from the perspective of distance functions, the projection point of the evaluated DMU is the point on the frontier farthest from that DMU, and this is the drawback and the unreasonableness of the SBM(ERM) model. From the standpoint of the evaluated unit, the frontier is expected to be reached along the shortest path, and the projected values supplied by the SBM(ERM) model clearly run counter to this.

The non-oriented SBM(ERM) model is a nonlinear programming model and can be converted into a linear programme by the following steps:

1) Let $\mathrm{t} = \dfrac{1}{1+\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$, and the model is(4.4) converted into

$$\min \ \rho = t - \frac{1}{m}\sum_{i=1}^{m} ts_i^- / x_{ik}$$

$$s.t. \ Xt\lambda + ts^- - tx_k = 0$$

$$Yt\lambda - ts^+ - ty_k = 0$$

$$\mathrm{t} = \frac{1}{1+\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.5)$$

2) Let $S^- = ts^-; S^+ = ts^+; \Lambda^- = t\lambda^-$, and the model is further converted into the following linear programme

$$\min \ \rho = t - \frac{1}{m}\sum_{i=1}^{m} S_i^- / x_{ik}$$

$$s.t. \ X\Lambda + S^- - tx_k = 0$$

$$Y\Lambda - S^+ - ty_k = 0$$

$$\mathrm{t} + \frac{1}{q}\sum_{r=1}^{q} S_r^+ / y_{rk} = 1$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.6)$$

In the objective function of the model(4.4), taking only the numerator or only the denominator yields the input-oriented and the output-oriented SBM(ERM) models respectively.

The input-oriented SBM(ERM) model is

$$\min \ \rho = 1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda \geq y_k$$

$$\lambda, s^- \geq 0 \quad (4.7)$$

In the input-oriented SBM(ERM) model, the input data may not contain zeros.

The output-oriented SBM(ERM) model is

$$\min \ \rho = \frac{1}{1 + \frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}} \text{ 或} \max \ \frac{1}{\rho} = 1 + \frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}$$

$$s.t. \ X\lambda \leq x_k$$

$$Y\lambda - s^+ = y_k$$

$$\lambda, s^+ \geq 0 \quad (4.8)$$

In the output-oriented SBM(ERM) model, the output data may not contain zeros.

The input-oriented SBM(ERM) model is equivalent to the Russell input-oriented efficiency measure proposed byRolf Färe & Knox Lovell (1978), but the formulation of the SBM(ERM) model is easier to understand.

Table 4-1 summarizes the programming formulations of the SBM(ERM) models under the three orientations.

The efficiency scores obtained from radial models and from the SBM(ERM) model always satisfy the following relationship: VRS efficiency score >= CRS efficiency score.

In radial models, the scale efficiency score can be calculated from the numerical relationship between the two: scale efficiency = CRS efficiency / VRS efficiency (SE = TE/PTE). Equivalently, the CRS efficiency score can be decomposed as follows: radial CRS efficiency = radial VRS efficiency * scale efficiency (TE = PTE * SE).

According to the definition of returns to scale, scale efficiency should be calculated with radial models. In the SBM(ERM) model and other distance function models, the ratio of the CRS efficiency score to the VRS efficiency score is called the "scale effect" (Scale Effect) in this book and in the MaxDEA software, so as to distinguish it from scale efficiency.

For the same evaluated DMU, the input- and output-oriented SBM(ERM) efficiency scores are both greater than or equal to the non-oriented SBM(ERM) efficiency score.

Table 4-1 Programming Formulations of the Various SBM(ERM) Models

	Input-oriented	Output-oriented	Non-oriented

	$\min \ \rho = 1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}$	$\max \ \frac{1}{\rho} = 1 + \frac{1}{s}\sum_{r=1}^{s} s_r^+ / y_{rk}$	$\min \ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1 + \frac{1}{s}\sum_{r=1}^{s} s_r^+ / y_{rk}}$
CRS	$s.t. \ X\lambda + s^- = x_k$ $Y\lambda \geq y_k$ $\lambda, s^- \geq 0$	$s.t. \ X\lambda \leq x_k$ $Y\lambda - s^+ = y_k$ $\lambda, s^+ \geq 0$	$s.t. \ X\lambda + s^- = x_k$ $Y\lambda - s^+ = y_k$ $\lambda, s^-, s^+ \geq 0$
VRS	$e\lambda = 1$		
NDRS	$e\lambda \geq 1$		
NIRS	$e\lambda \leq 1$		
GRS	$L_{\Sigma\lambda} \leq e\lambda \leq U_{\Sigma\lambda}$		

4.2.2 The Weighted SBM(ERM) Model

In the SBM(ERM) model, the individual input (output) measures are of equal importance in the efficiency measurement. According toKaoru Tone (2001), different weights can be assigned to the individual input and output measures on the basis of the SBM(ERM) model, so as to reflect differences in the importance of the measures. The weighting does not affect the efficiency status of the evaluated DMUs, that is, the DMUs rated efficient by the SBM(ERM) model and by the weighted SBM(ERM) model are the same, but the efficiency scores and projected values of inefficient DMUs change.

The non-oriented weighted SBM(ERM) model (weighted SBM(ERM), SBM(ERM) with preference) is expressed as

$$\min \ \rho = \frac{1 - \frac{1}{\sum_{i=1}^{m} w_i^I}\sum_{i=1}^{m} w_i^I s_i^- / x_{ik}}{1 + \frac{1}{\sum_{r=1}^{q} w_r^O}\sum_{r=1}^{q} w_r^O s_r^+ / y_{rk}}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.9)$$

wI and wO denote the weights of the input and output measures.

The input-oriented weighted SBM(ERM) model is expressed as

$$\min \ \rho = 1 - \frac{1}{\sum_{i=1}^{m} w_i^I} \sum_{i=1}^{m} w_i^I s_i^- / x_{ik}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda \geq y_k$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.10)$$

The output-oriented weighted SBM(ERM) model is expressed as

$$\min \ \rho = \frac{1}{1 + \frac{1}{\sum_{r=1}^{q} w_r^O} \sum_{r=1}^{q} w_r^O s_r^+ / y_{rk}}$$

$$s.t. \ X\lambda \leq x_k$$

$$Y\lambda - s^+ = y_k$$

$$\lambda, \ s^+ \geq 0 \quad (4.11)$$

4.2.3 The MSBM Model

MSBM (modified slack based measure) was proposed by(Sharp, Meng, & Liu, 2007) as a variant of the SBM(ERM) model; in its objective function, the maximum possible improvement of each input or output of the evaluated DMU (range of possible improvement) serves as the denominator for computing its degree of inefficiency, that is,

$$\min \ \rho = \frac{1 - \frac{1}{m} \sum_{i=1}^{m} s_i^- / R_{ik}^-}{1 + \frac{1}{q} \sum_{r=1}^{q} s_r^+ / R_{rk}^+}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0$$

$$R_{ik}^- = x_{ik} - \min(\mathrm{x}_i)$$

$$R_{rk}^+ = \max(\mathrm{y}_r) - y_{rk} \quad (4.12)$$

If $R_{ik}^- = 0$, the corresponding s_i^- / R_{ik}^- term is removed from the objective function; if $R_{rk}^+ = 0$, the corresponding s_r^+ / R_{rk}^+ term is removed from the objective function.

Compared with the SBM(ERM) model, the advantage of MSBM is that negative values are permitted in the input and output measures, but its drawback is that it can be applied only to VRS models and not to

CRS models.[i]

The MSBM model with weights applied to the input and output measures is expressed as

$$\min \ \rho = \frac{1-\frac{1}{\sum w_i^I}\sum_{i=1}^{m} w_i^I s_i^- / R_{ik}^-}{1+\frac{1}{\sum w_r^O}\sum_{r=1}^{q} w_r^O s_r^+ / R_{rk}^+}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.13)$$

4.2.4 A Worked Example: Solving the SBM(ERM) Model and the Weighted SBM(ERM) Model in MaxDEA

The procedure for solving the SBM(ERM) model with MaxDEA is basically the same as that for radial models; the differences are as follows:

1) On the distance interface, select "2) Farthest distance to the strong efficient frontier (ERM: Pastor, et al 1999; SBM: Tone 2001)". To run the MSBM model, select the "Adjusted SBM (Sharp et al 2007)" option.[ii]

2) The multiplier form of the SBM(ERM) model does not have as clear an interpretation as the multiplier form of radial models, so MaxDEA does not provide a multiplier model for SBM(ERM); the dual solution is, however, provided in the results of the envelopment model.

We again use the data in **Table 2-6** to compare the results of the radial model and the SBM(ERM) model.

The projection point of SBM(ERM) is the point on the frontier farthest from the evaluated DMU, so the efficiency scores calculated by the two models satisfy the following relationship:

SBM(ERM) efficiency score <= radial model efficiency score

In the results, apart from the six efficient DMUs, for which the two efficiency scores are equal, the efficiency scores obtained from the SBM(ERM) model are all lower than those obtained from the radial model.

The "distance" in the SBM(ERM) model refers to the average proportion by which the individual input (output) measures are improved; therefore, in the projected values obtained from the input- (output-)

[i]The MSBM model is designed for use with the VRS model. In the CRS and NIRS models the maximum possible improvement in the input measures of the DMU k under evaluation may exceed Rik, that is, sik>Rik, so that the objective function may become negative. The MSBM model is therefore not applicable to the CRS and NIRS models.

[ii]MSBM is applicable only to the VRS model; selecting another returns-to-scale type may produce abnormal results.

oriented SBM(ERM) model and from the input- (output-) oriented radial model, the improvement proportions of the individual inputs (outputs) satisfy the following relationship:

Average improvement proportion of the SBM(ERM) model >= average improvement proportion of the radial model

Table 4-2 Comparison of Efficiency Scores and Projected Values between the Radial Model and the SBM(ERM) Model (Input-Oriented CRS)[i]

	Radial model					SBM(ERM) model				
DMU	**Efficiency score**	**Number of beds**	**Number of health technical personnel**	**Number of outpatient visits**	**Number of inpatient admissions**	**Efficiency score**	**Number of beds**	**Number of health technical personnel**	**Number of outpatient visits**	**Number of inpatient admissions**
Anhui	0.944	133136	125338	74967434	4391516	0.933	140997	114824	63344221	4391516
Beijing	0.917	80363	87385	104340626	2095434	0.798	80363	87385	104340626	2095434
Fujian	1.000	89947	93898	72059496	3087618	1.000	89947	93898	72059496	3087618

Table 4-3 Comparison of Efficiency Scores and Projected Values between the SBM(ERM) Model and the Weighted SBM(ERM) Model (Non-Oriented CRS)

	SBM(ERM)					Weighted SBM(ERM)				
DMU	**Efficiency score**	**Number of beds**	**Number of health technical personnel**	**Number of outpatient visits**	**Number of inpatient admissions**	**Efficiency score**	**Number of beds**	**Number of health technical personnel**	**Number of outpatient visits**	**Number of inpatient admissions**
Anhui	0.7038	140997	132739	116649529	4391516	0.8696	140997	116123	83884325	4391516
Beijing	0.7164	87596	103025	104589875	2845602	0.6075	87596	102941	104340626	2846770
Fujian	1.0000	89947	93898	72059496	3087618	1.0000	89947	93898	72059496	3087618

A high average improvement proportion does not mean that the improvement proportion of every input (output) is high. Taking the input-oriented model as an example, the frontier of a DEA model is a convex polyhedron, so when any two points on the strong efficient frontier are compared, it cannot happen that one point has all its inputs lower (or higher) than the other. In the data of this example there are two inputs, and when any two points on the strong efficient frontier are compared, if one point has a lower value of the first input than the other, then its second input must be higher than that of the other. The results reflect this regularity: the projection point of the SBM(ERM) model and the projection point of the radial model are both projection points on the frontier (the two projection points may coincide), and their projected input values satisfy the following regularity: if the number of beds in the SBM(ERM) model is lower than the number of beds in the radial model, then the number of health technical personnel in the SBM(ERM) model must be higher than the number of health technical personnel in the radial model.

The procedure for solving the weighted SBM(ERM) model in MaxDEA consists in adding weight

[i]The original input and output data are given in **Table** 2-6.

settings to the settings of the SBM(ERM) model. Select "Set input-output weights" and set the weights. The weights of the input measures and those of the output measures are independent of each other; here the weights of the two input measures are set to 1 and 2, and the weights of the two output measures are set to 1 and 5. Setting the weights of the two input measures to 0.5 and 1 and those of the two output measures to 2 and 10 has the same effect. A comparison of the results of the non-oriented SBM(ERM) model and the weighted SBM(ERM) model is given in Table 4-3.

4.3 Minimum Distance to the Weak Efficient Frontier

4.3.1 Computing the Minimum Distance to the Weak Efficient Frontier

The minimum distance to the weak efficient frontier (MinDW) is the shortest distance between the evaluated DMU and the frontier, regardless of whether its projection point on the frontier is strongly or weakly efficient. This approach was first proposed in(A. Charnes, Rousseau, & Semple, 1996; Briec, 1999) and can be expressed as m+q linear programmes (m being the number of input measures and q the number of output measures):

$$\max \quad \beta_z, \quad z = 1, 2, ..., m+q$$

$$s.t. \quad \sum_{j=1}^{n} \lambda_j x_{ij} + \beta_z e_i \le x_{ik}, i = 1, 2, ...\mathrm{m}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - \beta_z e_r \ge y_{rk}, r = 1, 2, ...q$$

$$\lambda \ge 0$$

e_i and e_r are constants; in each programming formulation only one e equals 1 and all the others equal 0, that is,

$$e_i = 1 \, if \; i = z; e_i = 0 \, if \; i \ne z$$

$$e_r = 1 \, if \; r = z - m; e_i = 0 \, if \; r \ne z - m$$

The efficiency score of each model is expressed as $\theta_z^* = \dfrac{1 - \frac{1}{m} \sum_{i=1}^{m} \beta_z^* e_i / \mathrm{x}_{ik}}{1 + \frac{1}{q} \sum_{r=1}^{q} \beta_z^* e_r / y_{rk}}$

The efficiency score of the MinDW model is expressed as $\theta_{\max}^* = \max(\theta_Z^*, \; z = 1, 2, ..., m+q)$, and the largest efficiency score corresponds to the smallest β *, that is, to the minimum distance to the frontier.

In the input-oriented case there are m linear programmes:

$$\max \quad \beta_z, \quad z = 1,2,...,m$$

$$s.t. \quad \sum_{j=1}^{n} \lambda_j x_{ij} + \beta_z e_i \le x_{ik}, i = 1,2,...\mathrm{m}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rk}, r = 1,2,...q$$

$$\lambda \ge 0$$

In the output-oriented case there are q linear programmes:

$$\max \quad \beta_z, \quad z = 1,2,...,q$$

$$s.t. \quad \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ik}, i = 1,2,...\mathrm{m}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - \beta_z e_r \ge y_{rk}, r = 1,2,...q$$

$$\lambda \ge 0$$

The efficiency score of the MinDW model is never smaller than the efficiency score of a directional distance function model with any direction vector or of any other distance type (for example radial and SBM(ERM)). In other words, the efficiency score obtained from the MinDW model is the largest.

4.3.2 A Worked Example: Solving the MinDW Model in MaxDEA

To solve the MinDW model in MaxDEA, select "3) Minimum distance to the weak efficient frontier (Charnes, et al 1996)" on the distance interface.

Using Table 2-6, a comparison of the efficiency scores obtained from the radial model, the SBM(ERM) model and the MinDW model is given in Table 4-4. The common settings of the three models are output-oriented, VRS.

Table 4-4 Comparison of Efficiency Scores of Different Types of Distance Function Models

Region	SBM(ERM)	Radial	MinDW
Anhui	0.706253	0.956817	0.977932
Beijing	0.817982	0.917493	0.956972
Fujian	1.000000	1.000000	1.000000
Gansu	0.721133	0.847287	0.916353
Guangdong	1.000000	1.000000	1.000000
Guangxi	0.879912	0.996415	0.998204
Guizhou	0.699040	0.969756	0.984646
Hainan	0.725478	0.932287	0.964617
Hebei	0.646840	0.945152	0.971803
Henan	0.764488	0.969803	0.984670
Anhui	0.706253	0.956817	0.977932
…	…	…	…

4.4 Minimum Distance to the Strong Efficient Frontier

Note that the minimum distance to the strong efficient frontier model is a mixed integer linear programme, and solving it takes far longer than other distance types.

4.4.1 Computing the Non-Oriented Minimum Distance to the Strong Efficient Frontier

For ease of exposition, the minimum distance to the strong efficient frontier model is abbreviated as the MinDS model.

Since it is clearly unreasonable for the SBM(ERM) model to adopt the farthest projection point on the strong efficient frontier, it is natural to think that this drawback can be overcome by using the nearest point on the strong efficient frontier as the projection point.

The non-oriented minimum distance to the strong efficient frontier uses the nearest point on the strong efficient frontier as the projection point. The following text explains the idea behind and the computational method for the minimum distance to the strong efficient frontier.

If, in the SBM(ERM) model(4.4) the objective function is changed from minimization (min) to maximization (max), can the nearest projection point on the strong efficient frontier then be obtained?

$$\max \ \rho = \frac{1-\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.14)$$

Solving the model(4.14) shows that this simple idea is not feasible. The result of solving this model is inevitably that all slack variables equal 0, that the efficiency scores of all evaluated DMUs equal 1, and that the benchmark and projected value of every evaluated DMU are the DMU itself. This is because

$\lambda_k = 1, \lambda_j = 0 (j \neq k), s^- = 0, s^+ = 0$ is an optimal solution of the model(4.14) in question.

If, in the model(4.14) the reference set is restricted to efficient DMUs, would that be feasible? That is, a two-stage approach is adopted: in the first stage the SBM(ERM) model is solved(4.4) to identify all efficient DMUs, which are denoted as the DMU set $E = \{j \mid \rho_j = 1\}$; then in the second stage the following model is solved

$$\max \ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1 + \frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \ \sum_{j \in E} \lambda_j x_{ij} + s^- = x_{ik}$$

$$\sum_{j \in E} \lambda_j y_{rj} - s^+ = y_{rk}$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.15)$$

Taking Table 4-5 as an example, the model is solved.(4.15)

Table 4-5 Sample Data

DMU	x1	x2	y
A	10	40	10
B	15	25	10
C	32	24	16
D	48	16	16
E	24	48	16
F	54	27	18
G	50	60	20

In the first stage the SBM(ERM) model is solved(4.4), and the results are shown in Table 4-6. The efficient DMUs are A, B, C and D, and the inefficient DMUs are E, F and G.

Table 4-6 Sample Data SBM(ERM) Model(4.4) Analysis Results

DMU	Efficiency score (ρ)	Benchmark	Projected value x1	x2	y
A	1.000	A(1.000)	10.0	40.0	10.0
B	1.000	B(1.000)	15.0	25.0	10.0
C	1.000	C(1.000)	32.0	24.0	16.0
D	1.000	D(1.000)	48.0	16.0	16.0
E	0.917	B(1.600)	24.0	40.0	16.0
F	0.833	C(0.675); D(0.675)	54.0	27.0	21.6
G	0.650	C(1.563)	50.0	37.5	25.0

In the second stage, for the inefficient DMUs, the model is solved again(4.15), and the results are shown in Table 4-7 below.

Table 4-7 Sample Data Model(4.15) Analysis Results

DMU	Efficiency score (ρ)	Benchmark	Projected value x1	x2	y
E	1.000	A(0.800); B(0.533); D(0.167)	24.0	48.0	16.0
F	0.944	A(0.300); D(0.937)	48.0	27.0	18.0
G	0.833	A(1.333); D(0.417)	33.3	60.0	20.0

The following problem can be found in the results: E is an inefficient DMU, yet the model(4.15) wrongly concludes that E is an efficient DMU. If the model(4.4) is used to test the projection points of the three DMUs, all of them turn out to be inefficient. This shows that the model(4.15) cannot solve the problem of finding the nearest projection point on the frontier either. The reason for the error can be illustrated by the following figure (Figure 4-1). The model(4.15) restricts the reference set to the four efficient DMUs A, B, C and D; as can be seen from the figure, even under the constraint (Σλ=1), any point within the shaded area can be constructed by a linear combination of the four efficient DMUs. For example, point E can be constructed by a linear

combination of A, B and D (in the figure this can be understood as: A' is first constructed from A and B, and E is then constructed from A' and D), so in the model(4.15) the originally inefficient E is wrongly judged to be efficient. Similarly, in the model(4.15) F and G look for the nearest point within the shaded area as their projection point, and the nearest projection point lies on the line AD, which wrongly takes the projection point off the frontier.

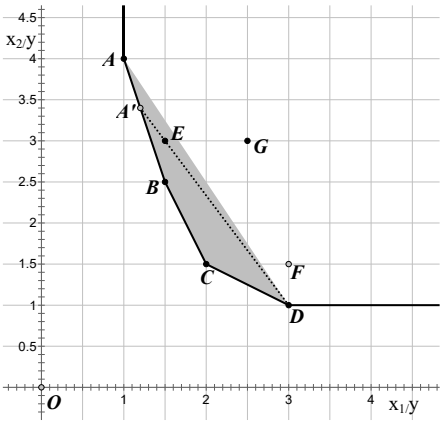


Figure 4-1 An Illustration of the Model(4.15) Error

The DEA frontier is composed of piecewise linear functions, each segment of which represents one facet (in a two-dimensional plane, a line segment); the model(4.15) produces this error because the reference set contains efficient DMUs lying on different facets. If the reference set in the model is restricted to efficient DMUs on the same facet, this error can be avoided. For example, in Figure 4-1 the frontier is divided into the three segments AB, BC and CD, and the model is then used(4.15) to calculate in turn the minimum distance from each inefficient DMU to each of the three frontier segments.

Following this line of reasoning, there are two methods for solving the MinDS model.

Method 1: find all supporting hyperplanes of the frontier (expressed as linear functions), then calculate separately the minimum distance from each inefficient DMU to each supporting hyperplane, and finally determine the smallest of these distances. Finding all supporting hyperplanes is a rather complicated computation; for the method see(Jahanshahloo, Lotfi, Rezai, & Balf, 2007; Jahanshahloo, Shirzadi, & Mirdehghan, 2009; F. H. Lotfi, Jahanshahloo, Mozaffari, & Gerami, 2011; Amirteimoori & Kordrostami, 2012), and the author of the SBM model, Tone, designed one method for solving the MinDS model along exactly

these lines(K. Tone, 2010).

Tone describes the computation of the MinDS model using the non-oriented model as an example:

1) Suppose there are n DMUs, and that the SBM(ERM) model(4.4) judges k of them efficient, the set of these k DMUs being denoted by $E=\{j \mid \rho_j = 1\}$ (these k DMUs are the vertices of all supporting hyperplanes of the frontier).

2) List all non-empty subsets of the set E. A new DMU is constructed by linear combination of all the DMUs in one subset (linear combination coefficients λ>0). For example, if the subset $\mathrm{E}_1=\{\mathrm{A}, \mathrm{B}\}$, the new DMU can be 0.5A+0.5B. This new DMU is placed into the original data and the SBM(ERM) model is used(4.4) to test whether it is an efficient DMU: if the new DMU is efficient, the subset is called an efficient subset of the set E. The DMUs in an efficient subset lie on the same supporting hyperplane of the frontier. The above process is repeated to find all efficient subsets of the set E.

3) Taking an efficient subset Ez of the set E as the reference set, the model is solved(4.16), and the efficiency score ρ_z is obtained. The difference between the model(4.16) and the model(4.15) is that in the model(4.15) the reference set is the set E consisting of all efficient DMUs, whereas in the model(4.16) the reference set is an efficient subset of the set E.

$$\max \ \rho = \frac{1-\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \ \sum_{j\in \mathrm{E}_z} \lambda_j x_{ij} + s^- = x_{ik}$$

$$\sum_{j\in \mathrm{E}_z} \lambda_j y_{rj} - s^+ = y_{rk}$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.16)$$

This process is repeated over all efficient subsets, and the efficiency score of the MinDS

model is

$$\rho^{\max} = \max\{\rho_z\},\ z = 1,2,...$$

When the number of efficient DMUs is large, the computation of the minimum distance by this method becomes extremely cumbersome. According to Tone's reasoning, in the CRS model there is no need to test subsets containing more than m+q-1 elements (m being the number of input measures and q the number of output measures); in the VRS model there is no need to test subsets containing more than m+q elements. If a subset is efficient, all subsets of this subset are certainly efficient; for example, if $\{A,B,C,D\}$ is efficient, then $\{A,B,C\}$ is certainly efficient; if a subset is inefficient, all subsets of this subset must be tested further, for example if $\{A,B,C,D\}$ is inefficient, then $\{B,C,D\}$, $\{A,C,D\}$, $\{A,B,D\}$ and $\{A,B,C\}$ must be tested to see whether they are efficient.[i] When the number of efficient DMUs is relatively large, the number of subsets to be tested becomes very large. Taking the CRS model as an example, when the number of inputs is 6 and the number of outputs is 5, if there are 20 efficient DMUs, the number of subsets to be tested is at least $C(20,10)=184756$[ii]; if there are 30 efficient DMUs, the total number of subsets to be tested is at least $C(30,10) = 30045015$. Since most subsets are inefficient, the number of tests actually required is much larger. This method is therefore applicable only when the number of efficient DMUs is small.

Method 2: adding constraints to the model(4.15) so that its benchmarks lie in the same hyperplane. Before Tone proposed the first method for solving the MinDS model,J. Aparicio, Ruiz, & Sirvent (2007) designed a simpler method. This method does not need to determine the hyperplanes of all frontiers; instead, by adding constraints it restricts the benchmarks of the DMU under evaluation to one and the same hyperplane. Once all efficient DMUs have been identified through the SBM(ERM) model, the method proposed by Aparicio needs only a single programming model to solve the MinDS model, regardless of how many efficient DMUs there are.

The methods of Aparicio and of Tone share the feature that both first obtain the set E of

[i]Among these subsets, if a subset is a subset of an efficient subset, it need not be tested again.

[ii]The number of such combinations can be calculated using the COMBIN function in Excel, $C(20,10)$ = COMBIN(20,10).

efficient DMUs through the SBM(ERM) model (or an equivalent model); they differ in that Aparicio et al. resort to mixed integer linear programming in order to avoid the cumbersome process of testing all subsets of E, so that it suffices, on the basis of the model(4.15), to add a set of mixed integer linear constraints, whereby the programming model automatically takes an efficient subset as its reference set and the testing of all subsets is neatly avoided. The steps of Aparicio's method for solving the MinDS model can be summarised as follows:

1) Suppose there are n DMUs, and let E be the set of DMUs judged efficient by the SBM(ERM) model;

2) solve the following mixed integer linear programme to obtain the MinDS efficiency score

$$\max \rho_k = \frac{\frac{1}{m}\sum_{i=1}^{m}(1-s_i^- / x_{ik})}{\frac{1}{q}\sum_{r=1}^{q}(1+s_r^+ / y_{rk})}$$

$$\text{s.t. } \sum_{j\in E} x_{ij}\lambda_j + s_i^- = x_{ik},\ i=1,2...,m \quad \text{(a-1)}$$

$$\sum_{j\in E} y_{rj}\lambda_j - s_r^+ = y_{rk}, r=1,2...,\mathrm{q} \quad \text{(a-2)}$$

$$s_i^- \geq 0,\ i=1,2...,m \quad \text{(a-3)}$$

$$s_r^+ \geq 0, r=1,2...,\mathrm{q} \quad \text{(a-4)}$$

$$\lambda_{\mathrm{j}} \geq 0,\ \mathrm{j}\in \mathrm{E} \quad \text{(a-5)}$$

$$-\sum_{\mathrm{i}=1}^{\mathrm{m}} v_{\mathrm{i}}\mathrm{x}_{\mathrm{ij}} + \sum_{\mathrm{r}=1}^{\mathrm{q}} \mu_{\mathrm{r}}\mathrm{y}_{\mathrm{rj}} + \mathrm{d}_{\mathrm{j}} = 0,\ \mathrm{j}\in \mathrm{E} \quad \text{(b-1)}$$

$$v_{\mathrm{i}} \geq 1,\ \mathrm{i}=1,2...,\mathrm{m} \quad \text{(b-2)}$$

$$\mu_{\mathrm{r}} \geq 1, \mathrm{r}=1,2...,\mathrm{q} \quad \text{(b-3)}$$

$$\mathrm{d}_{\mathrm{j}} \leq \mathrm{M}\mathrm{b}_{\mathrm{j}},\ \mathrm{j}\in \mathrm{E} \quad \text{(c-1)}$$

$$\lambda_{\mathrm{j}} \leq \mathrm{M}(1-\mathrm{b}_{\mathrm{j}}),\ \mathrm{j}\in \mathrm{E} \quad \text{(c-2)}$$

$$\mathrm{b}_{\mathrm{j}} \in \{0,1\},\ \mathrm{j}\in \mathrm{E} \quad \text{(c-3)}$$

$$\mathrm{d}_{\mathrm{j}} \geq 0,\ \mathrm{j}\in \mathrm{E} \quad \text{(c-4)(4.17)}$$

where M is a sufficiently large positive number.

The MinDS model(4.17) consists of three parts: the first part is the model(4.15), that is, the

objective function and constraint a; the second part is constraint b; the third part is constraint c.

The common purpose of the second part (constraint b) and the third part (constraint c) is to place the benchmarks in the same hyperplane. Although the model(4.17) does not solve for the hyperplane of the projected point, that is, this hyperplane is unknown, it is nevertheless possible, by adding constraints, to place the benchmarks within this unknown hyperplane.

The dual model (multiplier model) can be used to solve for the hyperplane in which the projected point of the DMU under evaluation and all of its benchmarks lie. Following this line of reasoning, additional constraints can be imposed on the multiplier model so that, on the basis that the reference set consists of efficient DMUs, the benchmarks are further restricted to the same hyperplane, thereby avoiding the problem that the benchmarks in the model(4.15) do not lie in the same hyperplane.

Since the additive model is simpler than the SBM(ERM) model and requires no linear transformation, the multiplier form of the additive model (constraint b) is adopted.

Although the optimal solution of the model(4.15) may lie off the frontier, among all feasible solutions of the model(4.15) there must exist a solution whose projected point is the point on the strong efficient frontier closest to the DMU under evaluation. This solution is denoted by S^k_{MinDSF}, and the set of its benchmarks is denoted by R^k_{MinDSF}. Since the additive model (based on slacks) or the weighted additive model is equivalent to the SBM(ERM) model in determining the set of efficient DMUs, it follows that S^k_{MinDSF} corresponds to a projected point that must also lie on the frontier of the additive model, that is, S^k_{MinDSF} is also a feasible solution of the additive model.

The additive model can be expressed as

$$\min \ -\sum_{i=1}^{m} s_i^- - \sum_{r=1}^{s} s_r^+$$

$$\text{s.t.} \ -\sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = -x_{ik}$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{rk}$$

$$\lambda, s^-, s^+ \geq 0$$

$$i = 1,2...,m; \ r = 1,2...,\mathrm{q}; \ \mathrm{j} = 1,2...,\mathrm{n} \quad (4.18)$$

Its dual model (multiplier form) is

$$\max -\sum_{i=1}^{m} v_i x_{ik} + \sum_{r=1}^{q} \mu_r y_{rk}$$

$$s.t. \ -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{q} \mu_r y_{rj} \le 0$$

$$v_i \ge 1$$

$$\mu_r \ge 1$$

$$i = 1, 2..., m; \ r = 1, 2..., q; \ j = 1, 2..., n \quad (4.19)$$

In the model(4.15) the feasible solution of minimum distance S^k_{MinDSF} is also(4.18) a feasible solution of the additive model. Suppose that the dual solution corresponding to this feasible solution in the dual model(4.19) is μ* and ν*; then the hyperplane in which all benchmarks of this feasible solution (R^k_{MinDSF}) lie can be expressed as[i]

$$-\sum_{i=1}^{m} v_i^* x_i + \sum_{r=1}^{q} \mu_r^* y_r = 0 .$$

A simple data set in Table 4-8 is used below to demonstrate this rule intuitively.

Table 4-8 Example Data for the Hyperplane

DMU	x1	x2	y
A	1.2	4.0	1.0
B	1.5	2.5	1.0
C	2.0	1.5	1.0
D	3.0	1.0	1.0
E	2.0	3.5	1.0
F	3.0	1.5	1.0
G	2.5	3.0	1.0

In Figure 4-2, the projection of G on the frontier at minimum distance is G', its benchmarks are A and B, and the hyperplane passing through A and B is $-5x_1 - x_2 + 10y = 0$ (the dashed line

[i]There may be more than one hyperplane passing through the projected point.

in the figure is the straight line d, which is the hyperplane when y=1). Apart from A and B, all other DMUs lie to the right of the hyperplane, that is, $-5x_1 - x_2 + 10y < 0$[i].

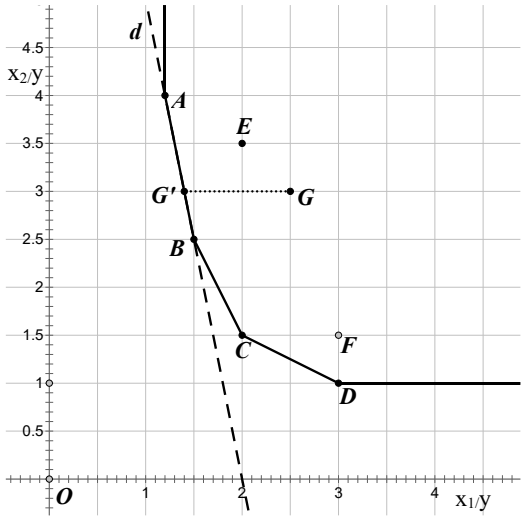


Figure 4-2 Illustration of the Hyperplane

From the above analysis it follows that, in the model(4.15): if DMU j lies in this hyperplane, then it should satisfy

$$-\sum_{i=1}^{m} v_i x_i + \sum_{r=1}^{q} \mu_r y_r = 0,\ j \in E .$$

The two sets of constraints added to the model(4.17) serve precisely to satisfy this rule:

1) if bj=0, then dj=0, and constraint (b-1) is equivalent to $-\sum_{i=1}^{m} v_i x_i + \sum_{r=1}^{q} \mu_r y_r = 0,\ j \in E$;

constraint (c-2) is equivalent to $\lambda_j <= M$, which, provided that M is sufficiently large, amounts to imposing no upper bound on λj. In this case DMUj is a benchmark for DMUk.

2) if bj=1, then dj<=M, which, provided that M is sufficiently large, amounts to imposing no

[i]In addition to A and B, if other DMUs also lie in this hyperplane, for example if there is a point H on the line joining A and B, then the model has multiple optimal solutions. The projected point G' may be constructed from {A, B}, {A, H} {B, H} or {A, B, H} in linear combination, and point H also satisfies $-5x_1 - x_2 + 10y = 0$, so in the model(4.17) the constraint imposed on DMUs that are not benchmarks is $-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{q} \mu_r y_{rj} \le 0$, rather than $-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{q} \mu_r y_{rj} < 0$.

upper bound on dj, and constraint (b-1) is equivalent to $-\sum_{i=1}^{m} v_i x_i + \sum_{r=1}^{q} \mu_r y_r \le 0, j \in E$; constraint (c-2) is equivalent to λj=0. In this case DMUj is not a benchmark for DMUk.

Using the model(4.17) the linear programming formulation of the MinDS model for E in Table 4-8 is as follows (with the big M value equal to 1000000); the solution results are reported in Table 4-9.

/* DMU_E */

/* Objective function */

max: +t -0.25 SLACK_x1 -0.142857142857 SLACK_x2;

/* Constraints */

Constraint_t: +t +SLACK_y = 1;

Constraint_x1: -2 t +1.2 LAMBDA_A +1.5 LAMBDA_B +2 LAMBDA_C +3 LAMBDA_D +SLACK_x1 = 0;

Constraint_x2: -3.5 t +4 LAMBDA_A +2.5 LAMBDA_B +1.5 LAMBDA_C +LAMBDA_D +SLACK_x2 = 0;

Constraint_y: -t +LAMBDA_A +LAMBDA_B +LAMBDA_C +LAMBDA_D -SLACK_y = 0;

Constraint_v_u_A: -1.2 v_1 -4 v_2 +u_1 +d_1 = 0;

Constraint_v_u_B: -1.5 v_1 -2.5 v_2 +u_1 +d_2 = 0;

Constraint_v_u_C: -2 v_1 -1.5 v_2 +u_1 +d_3 = 0;

Constraint_v_u_D: -3 v_1 -v_2 +u_1 +d_4 = 0;

Constraint_d_b_A: +d_1 -1000000 b_1 <= 0;

Constraint_d_b_B: +d_2 -1000000 b_2 <= 0;

Constraint_d_b_C: +d_3 -1000000 b_3 <= 0;

Constraint_d_b_D: +d_4 -1000000 b_4 <= 0;

Constraint_lambda_b_A: +LAMBDA_A +1000000 b_1 <= 1000000;

Constraint_lambda_b_B: +LAMBDA_B +1000000 b_2 <= 1000000;

Constraint_lambda_b_C: +LAMBDA_C +1000000 b_3 <= 1000000;

Constraint_lambda_b_D: +LAMBDA_D +1000000 b_4 <= 1000000;

/* Variable bounds */

v_1 >= 1;

v_2 >= 1;

u_1 >= 1;

b_1 <= 1;

b_2 <= 1;

b_3 <= 1;

b_4 <= 1;

/* Integer definitions */

int b_1,b_2,b_3,b_4;

Table 4-9 Computation Results of the MinDS Model for Point E

Variable	Result
Objective (efficiency score)	0.825
t	1
SLACK_x1	0.7
SLACK_x2	0
SLACK_y	0
LAMBDA_A	0.6667
LAMBDA_B	0.3333
LAMBDA_C	0
LAMBDA_D	0
v_1	833333.3333
v_2	166666.6667
u_1	1666666.6667
d_1	0
d_2	0
d_3	250000
d_4	1000000
b_1	0
b_2	0
b_3	1
b_4	1

Adding(4.17) the constraint $\sum_{j\in E}\lambda_j = 1$ and the free variable μ_0 to the CRS MinDS model yields the VRS MinDS model:

$$\max \rho_k = \frac{\frac{1}{m}\sum_{i=1}^{m}(1-s_i^- / x_{ik})}{\frac{1}{q}\sum_{r=1}^{q}(1+s_r^+ / y_{rk})}$$

$$\text{s.t. } \sum_{j\in E} x_{ij}\lambda_j + s_i^- = x_{ik},\ i = 1,2...,m$$

$$\sum_{j\in E} y_{rj}\lambda_j - s_r^+ = y_{rk}, r = 1,2...,\mathrm{q}$$

$$\sum_{j\in E}\lambda_j = 1$$

$$s_i^- \geq 0,\ i = 1,2...,m$$

$$s_r^+ \geq 0, r = 1,2...,\mathrm{q}$$

$$\lambda_{\mathrm{j}} \geq 0,\ \mathrm{j}\in \mathrm{E}$$

$$-\sum_{\mathrm{i}=1}^{\mathrm{m}} v_{\mathrm{i}}\mathrm{x}_{\mathrm{ij}} + \sum_{\mathrm{r}=1}^{\mathrm{q}} \mu_{\mathrm{r}}\mathrm{y}_{\mathrm{rj}} + \mu_0 + \mathrm{d}_{\mathrm{j}} = 0,\ \mathrm{j}\in \mathrm{E}$$

$$v_{\mathrm{i}} \geq 1,\ \mathrm{i} = 1,2...,\mathrm{m}$$

$$\mu_{\mathrm{r}} \geq 1, \mathrm{r} = 1,2...,\mathrm{q}$$

$$\mathrm{d}_{\mathrm{j}} \leq \mathrm{Mb}_{\mathrm{j}},\ \mathrm{j}\in \mathrm{E}$$

$$\lambda_{\mathrm{j}} \leq \mathrm{M}(1-\mathrm{b}_{\mathrm{j}}),\ \mathrm{j}\in \mathrm{E}$$

$$\mathrm{b}_{\mathrm{j}} \in \{0,1\},\ \mathrm{j}\in \mathrm{E}$$

$$\mathrm{d}_{\mathrm{j}} \geq 0,\ \mathrm{j}\in \mathrm{E}$$

$$\mu_0 \text{ free} \quad (4.20)$$

It should be noted that, owing to the numerical precision of linear programming solvers, M has to be a suitably large positive number; otherwise the value obtained may not be the correct optimum. If M is too large or too small, incorrect values may result. The appropriate range of M differs between solvers and between data sets.

Although both Tone's method and Aparicio's method can solve the minimum distance function and their underlying ideas are identical, Aparicio's method is clearly simpler and more practical. MaxDEA adopts Aparicio's method.

4.4.2 A Worked Example: Solving the Non-Oriented MinDS Model in MaxDEA

The MinDS distance involves mixed integer linear programming. When the number of DMUs is large, the solution may take a very long time and the software may temporarily stop responding, so patience is required.

This section uses real data to illustrate the process and results of solving the MinDS model in MaxDEA. Table 4-10 presents some of the input and output measures of hospitals in each province (municipality and autonomous region) in 2004. A DEA model is built with the number of beds and the number of health technical personnel as input measures, and the number of outpatient and emergency visits and the number of admissions as output measures.

MaxDEA solves the non-oriented MinDS model by Method 2, that is, theJ. Aparicio et al. (2007) method.

Table 4-10 Selected Input and Output Measures of Hospitals by Region (2004)

Region	Number of beds	Number of health technical personnel	Number of outpatient and emergency visits	Number of admissions
Anhui	77773	84672	36106166	1632009
Beijing	69850	90656	56191348	987106
Fujian	54975	51611	37915799	1290919
Gansu	45123	42130	19116626	708323
Guangdong	144416	157375	169171460	3545978
Guangxi	64466	66521	41722576	1418468
Guizhou	43835	39839	17156975	820696
Hainan	13251	16277	7449266	249455
Hebei	114373	128076	48433864	2399541
Henan	145870	160899	70404035	2935932

Heilongjiang	98975	100341	31470419	1568560
Hubei	95186	107546	50361992	2036096
Hunan	100167	106542	36818263	1957330
Jilin	69039	78462	26756299	1082780
Jiangsu	122787	129843	84927158	2872465
Jiangxi	58750	61535	31554864	1266984
Liaoning	139635	136979	47253345	2068764
Inner Mongolia	48074	53483	18296934	792838
Ningxia	14087	14243	7323406	260350
Qinghai	13004	12382	4781871	254352
Shandong	163876	181441	79416181	3785327
Shanxi	79721	88699	28674172	1176485
Shaanxi	77798	76225	29845077	1331931
Shanghai	65114	70265	58799940	1141108
Sichuan	123995	112744	65595496	2578469
Tianjin	36645	45229	24549117	521743
Tibet	4238	4918	2091942	52818
Xinjiang	61192	60869	27318653	1316919
Yunnan	71170	63095	34037932	1421345
Zhejiang	102295	107777	87342404	2434372
Chongqing	43784	38191	23643216	823952

4.4.2.1 The CRS Model of Minimum Distance to the Strong Efficient Frontier

Software operations: import the data, set the input and output measures, then select to run the envelopment model[i]

1) for Distance, select "4) Minimum distance to the strong efficient frontier";

[i]Solving the MinDS model takes far longer than solving the SBM model, and MaxDEA temporarily stops responding while it runs, especially when the data set is large, so patience is required.

2) for Returns to Scale, select "1) Constant (CRS)".

Table 4-11 reports part of the output of MaxDEA. For the 29 inefficient DMUs, the MinDS efficiency scores are on average 0.15 higher than the SBM(ERM) efficiency scores.

Table 4-11 Comparison of MinDS and SBM(ERM) Results (Non-Oriented, CRS)

Region	SBM(ERM)		MinDS	
	Efficiency score	Benchmark	Efficiency score	Benchmark
Anhui	0.5418	Guangdong (0.538027)	0.7150	Fujian (1.414698)
Beijing	0.5760	Guangdong (0.483672)	0.6534	Fujian (0.967877); Guangdong (0.115229)
Fujian	1.0000	Fujian (1.000000)	1.0000	Fujian (1.000000)
Gansu	0.5006	Guangdong (0.267705)	0.6420	Fujian (0.816299)
Guangdong	1.0000	Guangdong (1.000000)	1.0000	Guangdong (1.000000)

4.4.2.2 The VRS Model of Minimum Distance to the Strong Efficient Frontier

In the Returns to Scale options of the MaxDEA envelopment model, select "2) Variable (VRS)"; all other settings are the same as for the CRS model.

Table 4-12 compares the CRS and VRS results of the MinDS model. As noted earlier, in radial models and the SBM(ERM) model the CRS efficiency score is less than or equal to the VRS efficiency score. No such relationship holds in the MinDS model. **Table 4-12** contains 15 DMUs whose CRS efficiency score exceeds their VRS efficiency score, and the scale effect of these 15 DMUs (the ratio of the CRS efficiency score to the VRS efficiency score) is greater than 1. This issue is analysed specifically in the next section.

Table 4-12 Comparison of MinDS CRS and VRS Results (Non-Oriented)

Region	VRS		CRS	
	Efficiency score	Benchmark	Efficiency score	Benchmark
Anhui	0.6459	Fujian (0.848745); Guangdong (0.151255)	0.7150	Fujian (1.414698)
Beijing	0.6463	Fujian (0.860764); Guangdong (0.139236)	0.6534	Fujian (0.967877); Guangdong (0.115229)
Fujian	1.0000	Fujian (1.000000)	1.0000	Fujian (1.000000)
Gansu	0.6954	Fujian (0.437956); Qinghai (0.562044)	0.6420	Fujian (0.816299)
Guangdong	1.0000	Guangdong (1.000000)	1.0000	Guangdong (1.000000)
Guangxi	0.8612	Fujian (0.943439); Guangdong (0.056561)	0.8954	Fujian (1.172642)
Guizhou	0.7149	Fujian (0.546365); Qinghai (0.453635)	0.6740	Fujian (0.771909)
Hainan	0.8663	Fujian (0.127950); Qinghai (0.287609); Tibet (0.584440)	0.7131	Fujian (0.241037)
Hebei	0.5374	Fujian (0.508384); Guangdong (0.491616)	0.6690	Fujian (2.080455)
Henan	0.5649	Guangdong (0.925283); Shandong (0.074717)	0.7132	Fujian (2.653388)
Qinghai	1.0000	Qinghai (1.000000)	0.6456	Fujian (0.236544)
Shandong	1.0000	Shandong (1.000000)	0.7574	Fujian (2.980919)
Tibet	1.0000	Tibet (1.000000)	0.5513	Fujian (0.077090)

4.4.3 Peculiarities of the MinDS Model

4.4.3.1 The MinDS CRS Efficiency Score May Exceed the VRS Efficiency Score

The efficiency scores obtained from radial models, the SBM(ERM) model, and the directional distance function model and MinDW model introduced below all satisfy the following relationship:

VRS efficiency score >= CRS efficiency score.

Does this relationship still hold in the MinDS model? The existing literature does not discuss the question.

Table 4-13 lists the CRS and VRS efficiency scores of the MinDS model together with the scale effect. Of the 31 DMUs, 15 have a CRS efficiency score greater than their VRS efficiency score, and the scale effect calculated on this basis is greater than 1 (shown in bold in the table).

Table 4-13 Scale Effect of the MinDS Model

Region	CRS efficiency score	VRS efficiency score	Scale effect
Anhui	**0.7150**	**0.6459**	**1.1070**
Beijing	**0.6534**	**0.6463**	**1.0111**
Fujian	1.0000	1.0000	1.0000
Gansu	0.6420	0.6954	0.9232
Guangdong	1.0000	1.0000	1.0000
Guangxi	**0.8954**	**0.8612**	**1.0398**

This example shows that, in the MinDS model, the CRS efficiency score may exceed the VRS efficiency score. The objective function of the MinDS model maximises the efficiency score (minimises the distance). 1) Suppose that, compared with the MinDS CRS model, the MinDS VRS model merely adds the constraint $\Sigma\lambda=1$; the VRS efficiency score would then be smaller than the CRS efficiency score, because the optimal solution of the VRS model would in that case be a

feasible solution of the CRS model. 2) Suppose that the constraint Σλ=1 is removed from the MinDS VRS model; in the VRS-SBM(ERM) model the number of DMUs judged efficient is generally larger than in the CRS-SBM(ERM) model, so the reference set of the MinDS VRS model contains more DMUs than that of the MinDS CRS model (the VRS reference set contains the CRS reference set), and the efficiency score of the MinDS VRS model will therefore be greater than that of the MinDS CRS model. In summary, the difference between the MinDS VRS model and the MinDS CRS model contains both a component that makes the CRS efficiency score larger and a component that makes the VRS efficiency score larger, so both outcomes can occur.

4.4.3.2 Expansion of the Reference Set May Increase the MinDS Efficiency Score

Table 4-14 Example Data

DMU	x1	x2	y
A	1.2	4	1
B	1.5	2.5	1
C	2	3.5	1
D	2	1.5	1
E	3	1	1
F	3	1.5	1
K	2.5	3	1

Suppose there are three sets of DMUs, S1, S2 and S3, for example

$$S_1=\{A, B, C, K\}, S_2=\{D, E, F, K\}$$

$$S_3=S_1 \cup S_2 = \{A, B, C, D, E, F, K\}$$

Ek1, Ek2 and Ek3 denote the efficiency scores of DMU K calculated from DEA models built with S1, S2 and S3 respectively as the reference set.

The efficiency scores obtained from radial models, the SBM(ERM) model, and the directional distance function model and MinDW model introduced below all satisfy the following relationship

$E_{k3} \le E_{k1}, E_{k3} \le E_{k2}$.

The MinDS model alone does not satisfy this relationship: when the reference set is expanded, the efficiency score obtained from the MinDS model may remain unchanged, decrease or increase.

Once the reference set is expanded, the frontier may change in several ways, for example:

1) the frontier does not change, in which case the efficiency scores calculated by all distance functions remain unchanged;

2) the frontier shifts outwards as a whole, that is, every part of the frontier moves outwards, in which case the efficiency scores obtained from all distance functions decrease;

3) the frontier does not move but becomes more complete, as shown in Figure 4-3: when the reference set is expanded from S1 to S3 or from S2 to S3, the frontier does not move but becomes more complete (extending from AB or DE to ABDE). In this case the efficiency score obtained from the MinDS model may remain unchanged or may increase, whereas the efficiency scores obtained from the other distance function models remain unchanged or decrease. In the MinDS model, when the reference set is expanded from S1 to S3, the projection and the efficiency score of K are unchanged, the projection point being K1 and the efficiency score 0.7800 in both cases; when the reference set is expanded from S2 to S3, the projection of K changes from K2 to K1 and the efficiency score increases from 0.7083 to 0.7800.

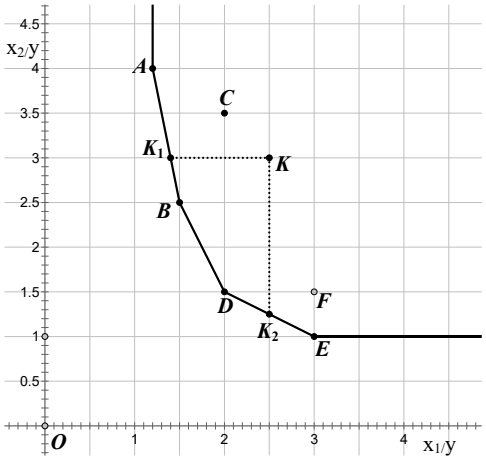


Figure 4-3 Illustration of the Increase in the Efficiency Score after Expanding the Reference Set

The increase in the efficiency score of the MinDS model after the reference set is expanded,

that is, the reduction in the minimum distance to the strong efficient frontier, arises because, before the reference set is expanded, there is a gap in the strong efficient frontier within the projection region of the DMU under evaluation (for example,Figure 4-4 the projection region of K is the polygonal region KK1DK2); the segment K1D is only weakly efficient. When the reference set is expanded, the gap is partly or completely filled, and the minimum distance from the DMU under evaluation to this filled part of the frontier is smaller than the minimum distance to the original frontier, so that after the reference set is expanded the distance from the DMU under evaluation to the new frontier decreases.

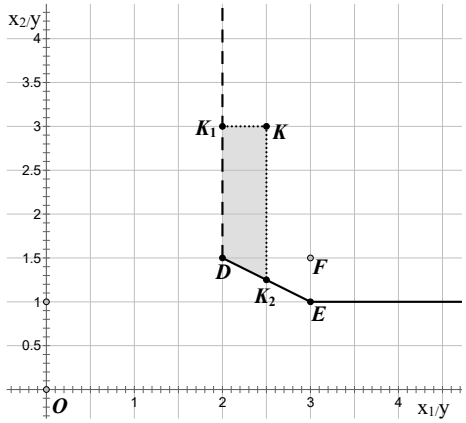


Figure 4-4 Illustration of the Cause of the Increase in the Efficiency Score after Expanding the Reference Set

For the reasons given above, anomalous results arise when MinDS is applied to the following models:

1) The sequential Malmquist model. One feature of the sequential Malmquist model is that the technological change index it produces is always greater than or equal to 1, that is, results obtained with the sequential Malmquist model never exhibit technological regress, which makes the model suitable for evaluating situations in which technological regress does not exist in theory(Shestalova, 2003). However, if the MinDS distance function is used in the sequential Malmquist model, anomalous results exhibiting technological regress may appear.

2) The meta-frontier model. In the meta-frontier model the technology gap ratio (TGR) equals the ratio of the meta-frontier efficiency score to the group-frontier efficiency score; since

the meta-frontier efficiency score is less than or equal to the group-frontier efficiency score, the resulting technology gap ratio is less than or equal to 1. However, if the MinDS distance function is used in the meta-frontier model, the meta-frontier efficiency score may exceed the group-frontier efficiency score, which produces the anomalous result of a technology gap ratio greater than 1.

4.4.3.3 The Monotonicity Problem

Another defect of MinDS is that it violates the monotonicity principle, that is, if the inputs of the DMU under evaluation decrease or its outputs increase, the resulting efficiency score may instead fall(Ando, Kai, Maeda, & Sekitani, 2012; Juan Aparicio & Pastor, 2013, 2014; Fukuyama, Maeda, Sekitani, & Shi, 2014; Ando, Minamide, Sekitani, & Shi, 2017).

4.4.3.4 Input- and Output-Oriented MinDS Models

The methods for solving the MinDS model designed by Tone and by Aparicio are both based on the non-oriented model. Although Tone states at the end of his paper that his method can equally be applied to input- and output-oriented models, this is not the case. Neither Tone's nor Aparicio's method for solving the MinDS model is applicable to input- or output-oriented models.

Taking the CRS model as an example, the programming formulation of the input-oriented MinDS model constructed with Tone's method is

$$\begin{aligned} &\max \rho_k^I = \tfrac{1}{m}\sum_{i=1}^{m}(1-s_i^- / x_{ik}) \\ &\text{s.t. } \sum_{j\in E_p} x_{ij}\lambda_j + s_i^- = x_k,\ i=1,2...,m \quad (4.21) \\ &\quad \sum_{j\in E_p} y_{rj}\lambda_j \geq y_k, r=1,2...,\mathrm{q} \end{aligned}$$

and the programming formulation of the input-oriented MinDS model constructed with Aparicio's method is

$$\begin{aligned} &\max \rho_k^I = \tfrac{1}{m}\sum_{i=1}^{m}(1-s_i^- / x_{ik}) \\ &\text{s.t. } \sum_{j\in E} x_{ij}\lambda_j + s_i^- = x_k,\ i=1,2...,m \\ &\quad \sum_{j\in E} y_{rj}\lambda_j \geq y_k, r=1,2...,\mathrm{q} \end{aligned}$$

$$\sum_{r=1}^{q}\mu_r y_{rj}-\sum_{i=1}^{m}\nu_i x_{ij}+d_j=0,\ j\in E$$

$$\nu_i\geq 1,\ \ i=1,2...,m$$

$$\mu_r\geq 0,\ r=1,2...,q$$

$$d_j\leq Mb_j,\ j\in E$$

$$\lambda_j\leq M(1-b_j),\ j\in E$$

$$b_j\in\{0,1\},\ j\in E$$

$$d_j\geq 0,\ j\in E$$

$$\lambda_j\geq 0,\ j\in E$$

$$s_i^-\geq 0,\ \ i=1,2...,m$$

$$s_r^+\geq 0,\ r=1,2...,q \quad (4.22)$$

Consider the simplest CRS model with a single input and a single output, and suppose there are three DMUs, A (x=4, y=1), B (x=3, y=3) and C (x=6, y=4), as shown in Figure 4-5. Solving the SBM(ERM) model(4.4) shows that B is efficient while A and C are inefficient. Then, using the model(4.21) to solve the MinDS model for A gives an efficiency score of 1, which contradicts the conclusion that A is inefficient. Figure 4-5 helps to explain why, in an input-oriented model, solving the MinDS model with the model(4.21) yields an erroneous result. In the input-oriented CRS model the objective function minimises the input distance between point A and the frontier, so that the projection of A on the frontier is AI; the input of AI equals that of A and the input slack of A is s-=0, so that A is misjudged as an efficient DMU. The frontier can be represented by the ray OB from the origin through point B, that is, x-y=0. Since x and y both take positive real values, every inefficient DMU has a projection point on the ray OB whose input equals its own, so that its input slack is zero.

Such a projection minimises input inefficiency at the expense of output inefficiency. Although the projection point is efficient, the measurement of the inefficiency of the DMU under evaluation in an input-oriented model ignores output inefficiency, so that the inefficiency of the

DMU under evaluation is shifted onto its outputs as far as possible. For example, in the figure, although the input of AI equals that of A, the output of AI is larger than that of A, the output slack of A being s+=3; this output slack is nevertheless not reflected in the objective function (the efficiency score).

In the input-oriented CRS model with a single input and a single output, an inefficient DMU can always attain the minimum value of the input slack by sacrificing output slack, and this minimum input slack is always zero. That is to say, in the single-input, single-output input-oriented model, if the MinDS model of an inefficient DMU is solved by Tone's or Aparicio's method, every inefficient DMU is wrongly judged to be efficient.

Similarly, in the output-oriented CRS model with a single input and a single output, an inefficient DMU can always attain the minimum value of the output slack by sacrificing input slack, and this minimum output slack is always zero. In Figure 4-5, under the output-oriented CRS model, the projection of A on the frontier is AO; the outputs of the two are identical, that is, the output slack of A is zero.

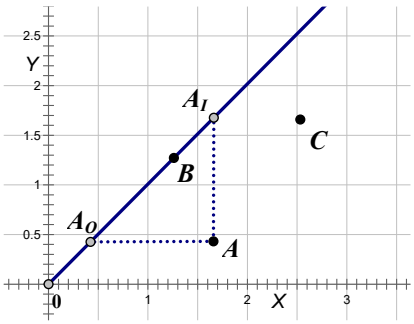


Figure 4-5 Example of the Minimum Distance Function with a Single Input and a Single Output

Consider next an example of a CRS model with two inputs and one output. As shown in Figure 4-6, points B and C are efficient DMUs. According to Tone's proof, every linear combination of points B and C (with positive coefficients) is efficient, that is, $E_1 = \{B,\ C\}$ constitutes an efficient subset. Suppose that B and C are taken as the reference set and that the DMU under evaluation is G; then the feasible region defined by Equation(4.21) is MBCNG. Since

the objective function is to be maximised, the optimal solution is G itself, that is, in the optimal solution determined by the programming formulation the linear combination of points B and C coincides with the DMU under evaluation (G) and the efficiency score is 1, which clearly contradicts the fact that G does not lie on the frontier. The same problem arises in the output-oriented model.

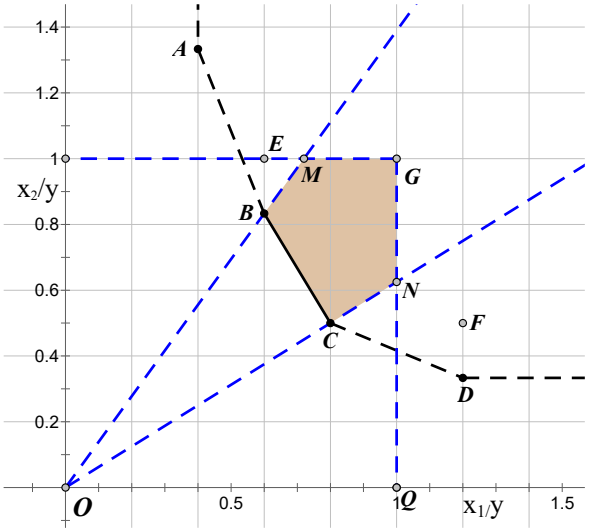


Figure 4-6 Example of an Erroneous Solution for the Maximum Efficiency Score in the Input-Oriented Model

In the VRS model, the same error arises when the MinDS model is solved by Tone's or Aparicio's method.

4.4.4 Recent Advances in Algorithms for the MinDS Model

The previous section explained that theJ. Aparicio et al. (2007) algorithm can be used only to solve the non-oriented MinDS model and cannot solve the input- or output-oriented MinDS models.Jahanshahloo, Vakili, & Zarepisheh (2012) proposed a theory and method for solving the minimum distance to the strong efficient frontier by means of bi-level programming, and pointed out that theJ. Aparicio et al. (2007) algorithm can be regarded as a special case of the bi-level programming approach. AlthoughJahanshahloo et al. (2012) did not directly give the programming formulation of the input- or output-oriented model, the formulations of the input- and output-oriented models can easily be derived from the non-oriented model they presented.J Aparicio, Cordero, & Pastor (2017) gave a bi-level programming algorithm for the output-oriented

model together with an equivalent single-level programming model (the algorithm for the input-oriented model is similar).

Two types of algorithm address the monotonicity problem.

The first is the weak monotonicity algorithm. This algorithm achieves weak monotonicity by extending the projection frontier in accordance with the principle of free disposability(Ando et al., 2012), for example theQ. Zhu, Wu, Ji, & Li (2018) weak monotonicity algorithm for the non-oriented model. Weak monotonicity means that, if the inputs of the DMU under evaluation decrease or its outputs increase, the resulting efficiency score does not fall. The radial distance efficiency score is weakly monotonic.

The second is the strong monotonicity algorithm. Strong monotonicity means that, if the inputs of the DMU under evaluation decrease or its outputs increase, the resulting efficiency score rises. This algorithm was proposed by Aparicio & Pastor(Juan Aparicio & Pastor, 2013, 2014), and adopts the Olesen & Petersen (1996) eFDEF (extended full dimensional efficient facet) as the projection frontier. There are two ways of implementing the strong monotonicity algorithm: one is to identify all FDEFs and then calculate the distance from an inefficient DMU to the eFDEF, for example theJuan Aparicio & Pastor (2013) method; the other is to adopt mixed integer (linear) programming (MIP) to realise the eFDEF as the projection frontier indirectly, for example theQ. Zhu, Aparicio, Li, Wu, & Kou (2022) algorithm.

The advantage of the eFDEF method is its speed, which allows it to be used on large data sets; its disadvantage is the complexity of the modelling. The advantage of the MIP method is the simplicity of the modelling; its disadvantages are slow speed and unsuitability for relatively large data sets.

Table 4-15 Comparison of Computation Times of the eFDEF and MIP Methods for the Output-Oriented Strongly Monotonic MinDS Model

eFDEF method		MIP method	
CRS	**VRS**	**CRS**	**VRS**
4 seconds	8 seconds	56 min 36 s	520 hours*

Test platform: Windows 11, Intel i9-9900K CPU with 8 cores at 3.6 GHz (CPU Turbo Boost disabled, that is, the operating frequency is fixed at 3.6 GHz).

Software: MaxDEA X 12.0, Gurobi 9.5. Computation uses 8 parallel threads.

Test data: uniformly distributed random data, 3 inputs and 3 outputs, 10,000 DMUs. The number of CRS-efficient DMUs is 61, and the number of VRS-efficient DMUs is 262.

*Because the running time is too long, this time is estimated from the average time taken to solve a single DMU.

4.5 Directional Distance Function

4.5.1 The Directional Distance Function Model

The directional distance function (DDF) model is a generalized expression of the radial DEA model (Chung, Färe, & Grosskopf, 1997).(Chung, Färe, & Grosskopf, 1997) In the directional distance function model, the researcher can define the direction in which the evaluated DMU is projected onto the frontier. In Euclidean space, the projection direction is determined by the direction vector, which consists of the input direction vector v and the output direction vector u; this is the most important feature of the directional vector model. Among the elements that make up the direction vector, the direction vector values of the different input and output measures represent their relative importance or priority. For example, if there are two inputs and the input vector is set to (2, 1), this means that, in projecting onto the frontier, the improvement of input 1 is given priority: for every 2 units of improvement in input 1, input 2 improves by 1 unit.

In Figure 4-7 An Example of the Input-Oriented Directional Distance Function Model, v1 and v2 define two direction vectors respectively, so that the projections of G onto the frontier are points B and C respectively. In theory, by defining the direction vector, an inefficient DMU can be projected onto the frontier along any specified direction.

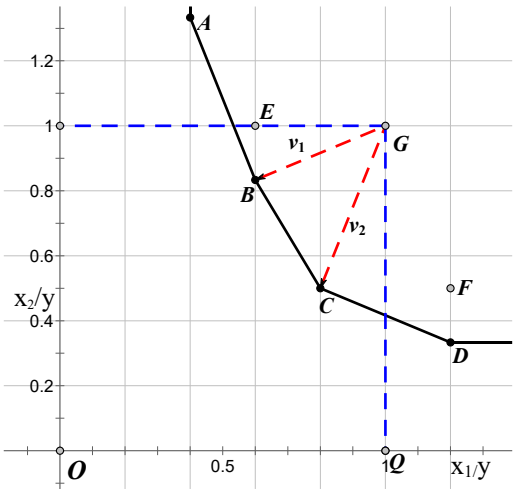


Figure 4-7 An Example of the Input-Oriented Directional Distance Function Model

Taking DMU k as the DMU under evaluation, the programming formulation of the directional vector model is

$$\max \beta$$

$$s.t.\ X\lambda + \beta g_x \le x_k$$

$$Y\lambda - \beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0 \quad (4.23)$$

In the model(4.23), $-g_x, g_y$ denotes the input-output direction vector, and the direction vector determines the direction in which the evaluated DMU is projected onto the frontier. The measurement of the degree of inefficiency (β) covers both the input and the output sides, so this is a non-oriented directional distance function model.

Model(4.23) is a CRS model; adding the constraint $\Sigma\ \lambda\ = 1$ to it yields the VRS model.

The result of the directional distance function (β) is independent of the units of measurement of the input and output measures; that is, for the same data set, as long as the direction vector is unchanged, the result of the model (β) remains unchanged when the units of measurement of the input and output measures change. It should be noted, however, that the direction vector values are values with units, and their units are the same as those of the corresponding input or output measures.[i] For example, in the sample data the units of the measures are bed, person, visit and person respectively; if the direction vector is written as (1, 1, 1, 1), the units of measurement are omitted, and the vector actually represents 1 bed, 1 person, 1 visit and 1 person. Suppose that the number of visits is measured in "10,000 visits" while the units of the other measures remain unchanged; then the direction vector value of the number of visits must also be expressed in "10,000 visits", and the revised direction vector is (1, 1, 0.0001, 1), which actually represents 1 bed, 1 person, 0.0001 ten-thousand visits and 1 person, that is, the direction vector values still represent 1 bed, 1 person, 1 visit and 1 person respectively. If the values of the direction vector are

[i] β is a coefficient without units, so the direction vector values must carry units, and those units must be consistent with those of the corresponding input or output measures; otherwise the two sides of the linear programming inequality will not be commensurate.

not adjusted in this way, the direction vector has in effect changed, and the analysis results of the model change accordingly.

As in the radial model, a slack problem also exists in the directional distance function model. The slack variable values can likewise be obtained by a two-stage approach, that is, by solving the following model in the second stage

$$\max \sum (s^- + s^+)$$

$$s.t.\ X\lambda + s^- = x_k - \beta^* g_x$$

$$Y\lambda - s^+ = y_k + \beta^* g_y$$

$$\lambda \geq 0$$

$$g_x \geq 0, g_y \geq 0 \quad (4.24)$$

4.5.2 The Multiplier Form of the Directional Distance Function Model

The CRS directional distance function model(4.23) has the following multiplier form (dual model)

$$\min \sum_{i=1}^{m} v_i x_{ik} - \sum_{r=1}^{s} \mu_r y_{rk}$$

$$s.t.\ \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0$$

$$\sum_{i=1}^{m} v_i g_i + \sum_{r=1}^{s} \mu_r g_r = 1$$

$$v \geq 0; \mu \geq 0$$

$$g_i \geq 0, g_r \geq 0 \quad (4.25)$$

The multiplier form of the directional distance function under the other types of returns to scale is

$$\min \sum_{i=1}^{m} v_i x_{ik} - \sum_{r=1}^{s} \mu_r y_{rk} + \mu_0$$

$$s.t. \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} + \mu_0 \geq 0$$

$$\sum_{i=1}^{m} v_i g_i + \sum_{r=1}^{s} \mu_r g_r = 1$$

$$v \geq 0; \mu \geq 0$$

VRS: μ_0 *free*

NIRS: $\mu_0 \geq 0$

NDRS: $\mu_0 \leq 0$

$$g_i \geq 0, g_r \geq 0 \quad (4.26)$$

4.5.3 Treatment of Undesirable Outputs in the Directional Distance Function Model and Whether Weak Disposability Is Reasonable

If undesirable outputs (also called bad outputs) exist, such as the pollution emissions generated in a production process, outputs can be distinguished within the directional distance function model: good outputs are denoted by Y and bad outputs by B, and the corresponding output vectors are divided into the good output vector gy and the bad output vector gb. The ability to treat good and bad outputs differently in the model is one of the main functions of the directional distance function model. After good and bad outputs are distinguished, the model is expressed as

$$\max \beta$$

$$s.t. \ X\lambda + \beta g_x \leq x_k$$

$$Y\lambda - \beta g_y \geq y_k$$

$$B\lambda + \beta g_b \leq b_k$$

$$g_x \geq 0, g_y \geq 0, g_b \geq 0 \quad (4.27)$$

$-g_x, g_y, -g_b$ are the direction vectors of inputs, good outputs and bad outputs, which

determine the direction in which the evaluated DMU is projected onto the frontier (reducing inputs, increasing good outputs and reducing bad outputs).

The production possibility set of the model(4.27) is

$$S = \{(\mathrm{x}, \mathrm{y}, \mathrm{b}) : \mathrm{x} \geq \mathrm{X}\lambda,\ y \leq Y\lambda,\ b \geq B\lambda\} \quad (4.28)$$

We use Table 4-16 to gain an intuitive understanding of the production possibility set of the model(4.27), where x is the input, y is the output and b is the undesirable output. In Figure 4-8, the production possibility set defined by the 7 DMUs is the shaded region OBCD that extends upward without limit along the vertical axis. This production possibility set implies that, given fixed inputs, bad outputs can increase without limit, that is, bad outputs exhibit strong disposability (strong disposable).

Table 4-16 Sample Data for the Directional Distance Function Model

DMU	x	y	b
A	1.0	2.0	1.0
B	1.0	3.5	1.5
C	1.0	4.2	2.0
D	1.0	4.5	2.5
E	1.0	1.8	2.0
F	1.0	2.0	2.5
G	1.0	3.0	4.0

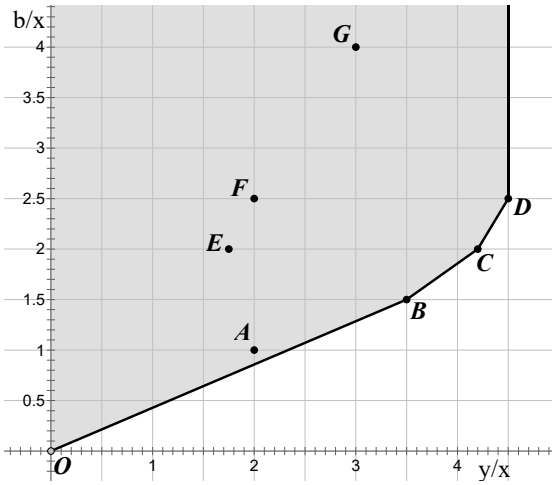


Figure 4-8 PPS of the Directional Distance Function with Strongly Disposable Undesirable Outputs

The production possibility set underlying the directional distance function model with

strongly disposable undesirable outputs is illogical. In actual production the quantity of bad outputs cannot increase without limit, and it should not exceed the maximum quantity that the production process is able to generate. Consider the production process of a coal-fired boiler for heating, which emits SO2. The quantity of SO2 emitted is determined by such factors as coal quality, the production process and waste gas treatment, and technological progress can gradually reduce emissions, which is the desired outcome. Conversely, in terms of production possibilities, using the lowest-quality coal and carrying out no waste gas treatment at all can maximize SO2 emissions, but this maximum is limited by the quantity of coal consumed (the input): given a fixed quantity of coal, SO2 emissions cannot increase without limit. The production possibility set of the directional distance function model with strongly disposable undesirable outputs is therefore logically flawed.

To avoid this problem,Chung et al. (1997) defined the production possibility set as follows in the directional distance function model they proposed:

1) Bad outputs are weakly disposable (weak disposability of bad outputs), which means that reducing bad outputs requires reducing good outputs as well, that is, given fixed inputs,

if $(y,b) \in P(x)$, then $(\theta y, \theta b) \in P(x), 0 \le \theta \le 1$ [i].

2) Good outputs are freely or strongly disposable (freely or strong disposability of good outputs), that is, given fixed inputs,

if $(y,b) \in P(x)$, then $(y\text{-s},b) \in P(x), s \ge 0$.

3) If bad outputs are 0, then good outputs are also 0, that is

if $(y,b) \in P(x)$ and b = 0, then y = 0.

The directional distance function with weakly disposable undesirable outputs is expressed as

[i] $P(x)$ denotes the output set, that is, $P(x)=\{(y,b) : \text{x 能生产}(y,b)\}$.

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda+\beta g_b = b_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0, g_b \ge 0 \quad (4.29)$$

After the weak disposability constraint is imposed on undesirable outputs, the production possibility set shrinks to the bounded region enclosed by OBCDGQ in Figure 4-9 (the shaded region).

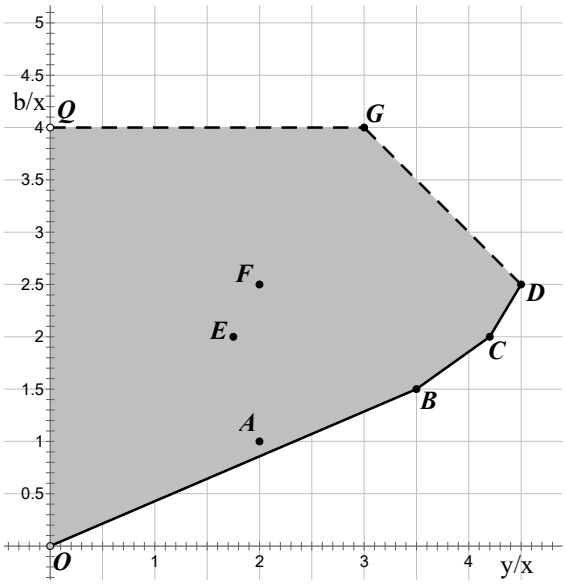


Figure 4-9 PPS of the Directional Distance Function with Weakly Disposable Undesirable Outputs

G is taken as an example to analyse the effect of weak disposability on the model. Among the 7 DMUs, G has the highest bad output but a lower good output than B, C and D. Suppose that G intends to improve its production efficiency through the following three options:

Option 1) Keep inputs and bad outputs unchanged and improve efficiency by increasing good outputs. To achieve this, the direction vector values for inputs, good outputs and bad outputs can be set to (0, 1, 0), and the corresponding linear programming formulation of the directional distance function is

$$\max \beta$$

$$s.t.\ X\lambda \le x_G$$

$$Y\lambda - \beta \ge y_G$$

$$B\lambda = b_G$$

$$\lambda \ge 0 \quad (4.30)$$

Option 2) Keep inputs unchanged and improve efficiency by reducing bad outputs while increasing good outputs, the ratio of the increase in good outputs to the decrease in bad outputs being 2:1. To achieve this, the direction vector values for inputs, good outputs and bad outputs can be set to (0, 2, -1), and the corresponding linear programming formulation of the directional distance function is[i]

$$\max \beta$$

$$s.t.\ X\lambda \le x_G$$

$$Y\lambda - 2\beta \ge y_G$$

$$B\lambda + \beta = b_G$$

$$\lambda \ge 0 \quad (4.31)$$

Option 3) Keep inputs unchanged and improve efficiency by reducing bad outputs while increasing good outputs, the ratio of the increase in good outputs to the decrease in bad outputs being 1:2. To achieve this, the direction vector values for inputs, good outputs and bad outputs can be set to (0, 1, -2), and the corresponding linear programming formulation of the directional distance function is

[i]The weak disposability of bad outputs introduced above (reducing a bad output necessarily requires reducing a good output at the same time) concerns the production possibility set of the production technology, that is, what is technologically (in terms of the production frontier) possible. The reduction of bad outputs with a simultaneous increase in good outputs discussed here refers to the way in which an inefficient DMU is improved, that is, the direction in which an inefficient DMU is projected towards the frontier. Taking pollution emissions as an example, the factors influencing a producer's decisions include not only internal production conditions but also external factors such as the severity of legal penalties for pollution emissions; the two should not be confused.

$$\max \beta$$

$$s.t.\ X\lambda \le x_{\mathrm{G}}$$

$$Y\lambda - \beta \ge y_G$$

$$B\lambda + 2\beta = \mathrm{b}_G$$

$$\lambda \ge 0 \quad (4.32)$$

The analysis results of the above three options are shown in **Table 4-17**. According to options 1 and 2, the projection of G is G itself, and only under option 3 is the projection a linear combination of B and C on the frontier.

Table 4-17 Analysis Results of the Directional Distance Function with Weakly Disposable Undesirable Outputs (G)

	β	Benchmark	x	y	b
Option 1	0.0000	G(1.0000)	1	3	4
Option 2	0.0000	G(1.0000)	1	3	4
Option 3	1.0526	B(0.2105); C(0.7895)	1	4.0526	1.8947

Constrained by the weak disposability of bad outputs, the production possibility set (the feasible region of the linear programme of the model) lacks the region to the lower right of G (the triangular region GDG2 in Figure 4-10), so G cannot be projected into this region.

Is the triangular region GDG2 of the production possibility set, which is lost because of the weak disposability constraint on bad outputs, in fact feasible in actual production?

Returning to the production process of a coal-fired boiler for heating: with the same quantity of inputs, G produces 3 units of good output and 4 units of bad output, whereas D produces 4.5 units of good output and 2.5 units of bad output. The production efficiency of D is better than that of G, mainly because of the improvement of its production process and its treatment of waste gas. In terms of production possibilities, by reducing its waste gas treatment while keeping good output unchanged, D could fully reach the level of bad output of G; that is, D is able to produce 4.5 units of good output while generating 4 units of bad output. This shows that part of the region of the production possibility set lost because of the weak disposability constraint on bad outputs should

in fact belong to the production possibility set.

Given fixed inputs and good outputs, it is technically possible to increase bad outputs within a certain range.

The weak disposability constraint on bad outputs creates another difficulty. On the face of it, the VRS model with weakly disposable bad outputs is obtained by adding the constraint $\Sigma\lambda = 1$ to the CRS model, that is

$$\begin{aligned}
&\max\ \beta \\
&s.t.\ X\lambda+\beta g_x \le x_k \\
&\quad Y\lambda-\beta g_y \ge y_k \\
&\quad B\lambda+\beta g_b = b_k \\
&\quad e\lambda=1 \\
&\quad \lambda \ge 0 \\
&\quad g_x \ge 0, g_y \ge 0, g_b \ge 0 \quad (4.33)
\end{aligned}$$

R. Färe & Grosskopf (2003) pointed out that the model(4.33) is not a correct VRS model with weakly disposable bad outputs, and that the correct model should be[i]

$$\begin{aligned}
&\max\ \beta \\
&s.t.\ X\lambda+\beta g_x \le x_k \\
&\quad \pi Y\lambda-\beta g_y \ge y_k \\
&\quad \pi B\lambda+\beta g_b = b_k \\
&\quad e\lambda=1 \\
&\quad 0 \le \pi \le 1 \\
&\quad \lambda \ge 0 \\
&\quad g_x \ge 0, g_y \ge 0, g_b \ge 0 \quad (4.34)
\end{aligned}$$

However, the model(4.34) has at least three problems: 1) it is a nonlinear programme, which

[i]In MaxDEA the VRS model with weak disposability of undesirable outputs uses the model(4.33), which should be used with caution.

is harder to solve than a linear programme; 2) its production possibility set is not a convex polyhedron; 3) it does not satisfy monotonicity, that is, an unreasonable result may arise in which the efficiency of the evaluated DMU rises after the DMU increases its bad outputs(Chen, 2013).

It is now worth returning to the question of whether removing the weak disposability constraint on bad outputs in the above three options would lead to erroneous results in the directional vector model. The linear programming models of the three options are respectively

Option 1)

$$\begin{aligned} &\max \ \beta \\ &s.t. \ X\lambda \le x_{\mathrm{G}} \\ &\quad Y\lambda - \beta \ge y_G \\ &\quad B\lambda \le \mathrm{b}_G \\ &\quad \lambda \ge 0 \end{aligned} \quad (4.35)$$

Option 2)

$$\begin{aligned} &\max \ \beta \\ &s.t. \ X\lambda \le x_{\mathrm{G}} \\ &\quad Y\lambda - 2\beta \ge y_G \\ &\quad B\lambda + \beta \le \mathrm{b}_G \\ &\quad \lambda \ge 0 \end{aligned} \quad (4.36)$$

Option 3)

$$\begin{aligned} &\max \ \beta \\ &s.t. \ X\lambda \le x_{\mathrm{G}} \\ &\quad Y\lambda - \beta \ge y_G \\ &\quad B\lambda + 2\beta \le \mathrm{b}_G \\ &\quad \lambda \ge 0 \end{aligned} \quad (4.37)$$

The analysis results are shown in **Table 4-18**. The results show that the projection values of

G are logical. The result of option 3 is the same as that of the weak disposability model, and the projection point under options 1 and 2 is point D in both cases. Why are the analysis results of the model not affected by its logically flawed production possibility set?

Table 4-18 Analysis Results of the Directional Distance Function with Strongly Disposable Undesirable Outputs (G)

	β	Benchmark	x	y	b
Option 1	1.5000	D(1.0000)	1	4.5	2.5
Option 2	0.7500	D(1.0000)	1	4.5	2.5
Option 3	1.0526	B(0.2105); C(0.7895)	1	4.0526	1.8947

In fact, in the directional distance function model with strongly disposable undesirable outputs, the evaluated DMU is usually projected in the direction of increasing good outputs and reducing bad outputs.[i] The projection value of bad outputs for any evaluated DMU is never greater than its original value, so the logically flawed production possibility set does not affect the projection of the evaluated DMU. Taking G as an example, in Figure 4-10, under the objective of increasing good outputs and reducing bad outputs (that is, the direction vector of good outputs is positive and that of bad outputs is negative), the projection range of G is the production possibility set region to its lower right, namely the polygonal region GG1BCDG2, while the region in which the production possibility set extends upward without limit owing to strongly disposable bad outputs never becomes a projection region of the evaluated DMU.

[i]An extreme case is one in which the quantity of bad outputs is disregarded altogether, for example when the government imposes no penalty at all on pollution emissions. In this case, from the producer's point of view, only an increase in good outputs is considered, and the model need not include bad outputs at all; good outputs alone suffice, and the question of whether bad outputs are strongly or weakly disposable no longer arises. Moreover, a rational decision would not seek to increase bad outputs without increasing good outputs. If the decision maker seeks to increase good and bad outputs simultaneously, that is, both direction vector values are positive, then in the programming formulation the direction of the inequality in the bad output constraint generally ensures that the optimal solution of the model does not exceed what the production technology can achieve, because in the optimal solution bad outputs increase together with good outputs in a certain proportion, and since good outputs cannot increase without limit, bad outputs cannot increase without limit either. The exception is when the direction vector values for good and bad outputs are set to extremely unreasonable magnitudes, for example gy=1 and gb=an extremely large unreasonable value.

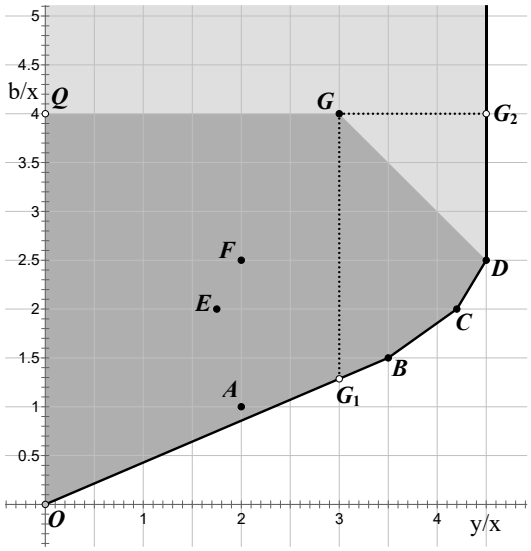


Figure 4-10 Projection of the Directional Distance Function with Strongly Disposable Undesirable Outputs

The above analysis leads the author to argue that imposing a weak disposability constraint on undesirable outputs in the directional distance function model, although widely accepted, is a specious treatment. In a word, although the PPS of the model with strongly disposable undesirable outputs is wrong, the results of the model are correct. It is recommended that, when undesirable outputs are present, no weak disposability constraint be imposed on them.

4.5.4 Some Special Cases of the Directional Distance Function Model

The directional distance function model is a generalization of the radial model. When the direction vector takes the input and output values of the evaluated DMU, the directional distance function model is identical to the radial DEA model; in this case the relationship between β* in the directional distance function model and the efficiency value θ* of the input-oriented radial DEA model is $\beta^* = 1 - \theta^*$, and its relationship with the efficiency value $1/\varphi^*$ of the output-oriented radial DEA model is $1/\varphi^* = 1/(1 + \beta^*)$.

When gx = xk and gy = 0, the directional distance function model is equivalent to the input-oriented radial model:

$$\max \beta \quad \text{or} \quad \min \theta=1-\beta$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或} X\lambda \le (1-\beta)x_k \text{ 或} X\lambda \le \theta x_k$$

$$Y\lambda \ge y_k$$

$$\lambda \ge 0 \quad (4.38)$$

When gx = 0 and gy = yk, the directional distance function model is equivalent to the output-oriented radial model:

$$\max \beta \quad \text{or} \quad \max \phi=1+\beta$$

$$s.t.\ X\lambda \le x_k$$

$$Y\lambda-\beta y_k \ge y_k \text{ 或} Y\lambda_k \ge (1+\beta)y_k \text{ 或} Y\lambda_k \ge \phi y_k$$

$$\lambda \ge 0 \quad (4.39)$$

When gx = xk and gy = yk, the directional distance function model is a form of the non-oriented radial model:

$$\max \beta \quad \text{or} \quad \min \frac{1-\beta}{1+\beta}$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或 } X\lambda \le (1-\beta)x_k$$

$$Y\lambda-\beta y_k \ge y_k \text{ 或} Y\lambda_k \ge (1+\beta)y_k$$

$$\lambda \ge 0 \quad (4.40)$$

When undesirable outputs are present, the three special cases above take the following forms.

When gx = xk, gy = 0 and gb = 0, the directional distance function model is equivalent to the input-oriented radial model with undesirable outputs:

$$\max \beta \quad \text{or} \quad \min \theta=1-\beta$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或} X\lambda \le (1-\beta)x_k \text{ 或} X\lambda \le \theta x_k$$

$$Y\lambda \ge y_k$$

$$B\lambda \le b_k$$

$$\lambda \ge 0 \quad (4.41)$$

When gx = 0, gy = yk and gb = bk, the directional distance function model is equivalent to the output-oriented radial model with undesirable outputs:

$$\max \beta \quad \text{or} \quad \min 1/\phi = \frac{1}{1+\beta}$$

$$s.t.\ X\lambda \le x_k$$

$$Y\lambda-\beta y_k \ge y_k \text{ 或} Y\lambda_k \ge (1+\beta)y_k \text{ 或} Y\lambda_k \ge \phi y_k$$

$$B\lambda+\beta b_k \le b_k \text{ 或} B\lambda \le (1-\beta)b_k$$

$$\lambda \ge 0 \quad (4.42)$$

When gx = xk, gy = yk and gb = bk, the directional distance function model is equivalent to the non-oriented radial model with undesirable outputs:

$$\max \beta \text{ 或} \min \frac{1-\beta}{1+\beta}$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或} X\lambda \le (1-\beta)x_k$$

$$Y\lambda-\beta y_k \ge y_k \text{ 或} Y\lambda_k \ge (1+\beta)y_k$$

$$B\lambda+\beta b_k \le b_k \text{ 或} B\lambda \le (1-\beta)b_k$$

$$\lambda \ge 0 \quad (4.43)$$

4.5.5 Computing the Efficiency Score of a General Directional Distance Function Model

Apart from the special cases above, β* cannot be used directly to represent the degree of inefficiency. β* is affected by the length of the direction vector: for example, in the directional distance function model of the same DMU, if β* = 0.5 when the direction vector is (1, 1,...), then β* = 0.25 when the direction vector is (2, 2,...).

$\beta \mathrm{g}_{\mathrm{x}_i}$ denotes the amount of improvement in each input, $\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}$ denotes the proportion of improvement in each input, $\frac{1}{m}\sum_{i=1}^{m}\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}$ denotes the average proportion of improvement in the inputs, which can be used to represent the degree of input inefficiency; $\beta \mathrm{g}_{\mathrm{y}_r}$ denotes the amount of improvement in the outputs, $\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}$ denotes the proportion of improvement in the outputs, $\frac{1}{q}\sum_{r=1}^{q}\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}$ denotes the average proportion of improvement in the outputs, which can be used to represent the degree of output inefficiency. The efficiency value of the directional distance function can therefore be defined as(Cheng & Zervopoulos, 2014):

$$\theta_k = \frac{1-\frac{1}{m}\sum_{i=1}^{m}\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q}\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}} \quad (4.44)$$

Because of the monotonic relationship between β and the objective function, the following two ways of computing the efficiency value give the same result:

1) First solve the directional distance function model to obtain the optimal solution β*, and then use β* to compute the efficiency value θ;

2) Use the formula for the efficiency value directly as the objective function of the directional distance function model, that is

$$\min\ \theta_k = \frac{1-\frac{1}{m}\sum_{i=1}^{m}\beta g_{x_i}/x_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q}\beta g_{y_r}/y_{rk}}$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0 \quad (4.45)$$

The efficiency value of the input-oriented directional distance function model is[i]

$$\min\ \theta_k = 1-\frac{1}{m}\sum_{i=1}^{m}\beta g_{x_i}/x_{ik}$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0 \quad (4.46)$$

The efficiency value of the output-oriented directional distance function model is

$$\min\ \theta_k = \frac{1}{1+\frac{1}{q}\sum_{r=1}^{q}\beta g_{y_r}/y_{rk}}$$

$$s.t.\ X\lambda \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_y \ge 0 \quad (4.47)$$

[i]The input-oriented directional distance function is equivalent to a non-oriented directional distance function in which the output direction vector value equals 0; the output-oriented directional distance function is equivalent to a non-oriented directional distance function in which the input direction vector value equals 0.

The definition of the efficiency value of the directional distance function(4.44) has the following features:

1) It is consistent (compatible) with the efficiency values of the radial DEA models: when the direction vector takes the three special forms equivalent to the radial models described above, the efficiency value obtained from the directional distance function model is the same as that of the radial model.

2) The efficiency value of the directional distance function obtained from the definition(4.44) is independent of the length of the direction vector used in the model. For example, in the same model, the efficiency values obtained are equal when the direction vector is the vector (1, 1,..., 1) with all elements equal to 1 and the vector (2, 2,..., 2) with all elements equal to 2.

3) With(4.44) as the basis, inputs and outputs can be readily weighted so as to reflect the importance of the different measures,

$$\theta_k = \frac{1-\frac{1}{\sum w_i^I}\sum_{i=1}^{m} w_i^I \beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{\sum w_r^O}\sum_{r=1}^{q} w_i^O \beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}} \quad (4.48)$$

when undesirable outputs are present,

$$\theta_k = \frac{1-\frac{1}{\sum w_i^I}\sum_{i=1}^{m} w_i^I \beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{\sum w_r^G+\sum w_t^B}\left(\sum_{r=1}^{q} w_r^G \beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}+\sum_{t=1}^{p} w_t^B \beta(-\mathrm{g}_{b_t}) / b_{tk}\right)} \quad (4.49)$$

4.5.6 A Worked Example: Solving the Directional Distance Function Model in MaxDEA

4.5.6.1 An Example of the Directional Distance Function Model

The **Table 2-6** are used as an example to illustrate the process of solving a general directional distance function model in MaxDEA.

First, in the distance interface, select "5) Directional Distance Function (Chambers, et al 1996; Chung, et al 1997)".

In the directional distance function model, the types of direction vector provided by MaxDEA include:

a): (-|x0|, |y0|, -|b0|)': this option is the special case described above, in which the directional distance function is equivalent to the radial model, with gx, gy and gb taking the absolute values of the input and output values of the evaluated DMU.

b): (-|$\bar{x}$|, |$\bar{y}$|, -|$\bar{b}$|)': gx, gy and gb take the absolute values of the means (averages) of the input and output measures.

c): (-1, 1, -1)': gx, gy and gb all equal 1, which means that the direction vector is $\begin{bmatrix} -1 \\ ... \\ 1 \\ ... \\ -1 \end{bmatrix}$.

d): Range (RDM, Portela, Thanassoulis, and Simpson 2004): this is a method proposed byPortela, Thanassoulis, & Simpson (2004) in which the maximum possible improvement of the evaluated DMU (range of possible improvement) is used as the direction vector value (range directional model, RDM), that is

$$g_{xk} = x_k - \min(\mathrm{x})$$

$$g_{yk} = \max(\mathrm{y}) - y_k$$

AlthoughPortela et al. (2004) did not give a direction vector for undesirable outputs directly, the direction vector value can be defined, in accordance with the direction of improvement of undesirable outputs, as $g_{bk} = b_k - \min(b)$.

e) Customized (the same vector for all DMUs): the direction vector is defined by the user and the defined vector is used for all DMUs. Click the "Settings" button and enter the direction vector values of the input and output measures; the signs need not be considered when entering them. Enter 1, 1, 1 and 15, which correspond to the number of beds, the number of health technical personnel, the number of visits and the number of inpatients respectively.

f) Customized (a different vector for each DMU): a direction vector value is defined separately for each evaluated DMU, that is, a different direction vector is used when different DMUs are evaluated. To use this type of direction vector, variables for the direction vectors of the individual input and output measures must be added to the data before the data are imported (for an example, see Table 4-19); as many columns of vector values are added as there are input and output measures. These added measures are kept as "Not defined" when the data are defined. Then, in the Define interface, specify the vector variable of each input and output.

When customized direction vectors (e and f) are entered, the signs of the values need not be considered: MaxDEA ignores the sign of the values entered by the user and determines the sign to be used automatically.

Table 4-19 Example of Adding Direction Vector Measures to the Data

DMU	x	y	x vector values	y vector values
A	88	81	1	11
B	57	70	2	10
C	57	75	2	11
D	68	90	1	10
E	58	83	1	11
F	82	69	2	10
G	57	73	2	11

All five preceding types of direction vector can be implemented through type 6. For example, the maximum possible improvement of each input and each output can be computed before the data are imported and added to the original data as the direction vector variables of the respective measures. Selecting type 6 of direction vector is then equivalent to selecting type 4 directly, that is, Range (RDM: Portela et al, 2004).

This example selects type 5 of direction vector, that is, customized (the same vector for all

DMUs), and enters gx as (1, 1) and gy as (1, 15),[i] which means that the direction vector is $\begin{bmatrix} -1 \\ -1 \\ 1 \\ 15 \end{bmatrix}$.

Note:

1) To separate the direction vectors of inputs, good outputs and bad outputs, select the last option, "8) Non-oriented (customized priority)", in the orientation interface.

2) Choosing input orientation is equivalent to choosing non-orientation and setting the output direction vector values to 0 through customization. Under input orientation the output direction vector values are not used in the model, so they do not affect the results.

3) Choosing output orientation is equivalent to choosing non-orientation and setting the input direction vector values to 0 through customization. Under output orientation the input direction vector values are not used in the model, so they do not affect the results.

In this example the model orientation is Output-oriented, and the returns to scale are CRS.

If weights are to be set for the measures, select "Preference (weighting of inputs and outputs)" in the Advanced Models interface and set the weights.[ii]

In an input-oriented model the weights of output measures have no effect; in an output-oriented model the weights of input measures have no effect. No measure weights are set in this example, which is equivalent to giving all measures the same weight.

The analysis results are shown in **Table 4-20**. When all direction vector values are greater than 0, the directional distance function model and the radial model are equivalent in judging whether the evaluated DMU is in a (weakly) efficient state, but the efficiency values of inefficient DMUs and their benchmarks differ. There is no fixed quantitative relationship between the efficiency values of the directional distance function model and those of the radial model; in this example the efficiency values of most DMUs under the directional distance function model are higher than under the radial model.

[i]The direction vector values set here serve demonstration purposes only and have no practical meaning.

[ii]Under some settings, for example when the first type of direction vector (the values of the DMU itself) is used, whether or not input and output weights are set has no effect whatsoever on the model; in this case the software displays the option "preference type (weighting inputs and outputs)" as unavailable (greyed out).

Table 4-20 Comparison of the Analysis Results of the Radial Model and the Directional Distance Function Model (Output-Oriented CRS)

Region	Directional distance function model		Radial model	
	Efficiency value	Benchmark	Efficiency value	Benchmark
Anhui	0.9712	Fujian (0.673719); Xinjiang (0.825137)	0.9442	Fujian (0.673719); Xinjiang (0.825137)
Beijing	0.7960	Fujian (0.005235); Guangdong (0.354095)	0.9174	Shanghai (1.000548)
Fujian	1.0000	Fujian (1.000000)	1.0000	Fujian (1.000000)

4.5.6.2 An Example of the Directional Distance Function Model with Undesirable Outputs

The following table shows, for the hospitals of each province (municipality, autonomous region) in 2010, the number of beds, the number of health technical personnel, the number of visits, the number of discharged patients and the number of cases in which the diagnosis at admission and the diagnosis at discharge are inconsistent. The number of beds and the number of health technical personnel are taken as inputs, the number of visits and the number of discharged patients as good outputs, and the number of inconsistent diagnoses as a bad output. When the data are defined (Define Data), the bad output measure is defined in the same way as the good outputs, namely as "Output".

Table 4-21 Selected Input, Output and Quality Measures of Hospitals by Region (2010)

Region	Number of beds	Number of health technical personnel	Number of visits	Number of discharged patients	Number of inconsistent admission and discharge diagnoses
Anhui	123427	123974	56090890	3717343	26021
Beijing	85775	121424	93376029	1708495	8542
Fujian	80896	82726	65632537	2710285	51495
Gansu	63773	50119	26462412	1484519	25237
Guangdong	224114	262214	257480147	7087548	42525
Guangxi	88913	97675	60570961	2910575	43659
Guizhou	69343	61260	25368195	2089306	37608
Hainan	18807	23324	11650197	528349	12152
Hebei	172956	175243	74748831	5099328	66291
Henan	220974	212847	103470565	6203957	111671
Heilongjiang	123928	122477	45083824	2721994	24498
Hubei	135006	138213	77101473	4232467	76184
Hunan	150141	141921	58694543	4684786	84326
Jilin	89341	83764	35533691	1994079	23929
Jiangsu	195340	195589	150420724	5577559	33465
Jiangxi	77805	85995	42630399	2524640	37870
Liaoning	160894	152822	66522624	3640376	32763
Inner Mongolia	67016	69118	28274251	1566913	29771
Ningxia	20258	20479	11314914	555371	3888
Qinghai	16226	15519	8135941	406122	6092
Shandong	255764	258707	123570567	7533675	75337
Shanxi	108260	115090	36443520	2132218	17058
Shaanxi	104819	111967	49244614	2849597	31346
Shanghai	84825	93566	104898763	2114263	10571
Sichuan	184828	172381	98207483	5557053	83356
Tianjin	40387	50897	43334154	933063	9331
Tibet	5444	5435	3390468	108864	2177
Xinjiang	89871	78206	34160068	2644559	29090
Yunnan	112493	83687	57102579	3316600	89548
Zhejiang	150986	171461	156681377	4345922	65189
Chongqing	64827	60294	34036573	1799571	23394

When undesirable outputs are present, the distance, the orientation, returns to scale and the measure weights are set in the same way as in the general directional distance function model.

This example model uses the first direction vector type, namely "a) = (-x0, y0, -b0)"; the orientation is set to "3) Non-oriented"; returns to scale is set to VRS. In addition, on the "Advanced Models" interface, the "Undesirable Output" option should be selected[i].

If the direction vector value of an output measure is set to a negative number through a custom direction vector (the fifth and sixth direction vector types) without that measure being defined as a bad output, the software in fact computes an incorrect model(4.27).

If the undesirable output is to be treated as weakly disposable, select the "Weak Disposability" option on the "Advanced Models" interface and set the undesirable output as weakly disposable.

This example model sets the undesirable output as strongly disposable (the "Weak Disposability" option is not selected) and as weakly disposable respectively, and compares the two sets of results[ii].

The analysis results are reported in Table 4-22. Since the optimal solution of the model in which the bad output is weakly disposable is a feasible solution of the strongly disposable model, the efficiency score of the weakly disposable model is greater than or equal to that of the strongly disposable model. In this example, the efficiency score of the weakly disposable model exceeds that of the strongly disposable model for 5 DMUs, and 3 of these DMUs (Henan, Hubei and Zhejiang) are inefficient in the strongly disposable model but efficient in the weakly disposable model.

Table 4-22 Analysis Results of the Directional Distance Function Model with Undesirable Outputs

Region	Strongly Disposable Model		Weakly Disposable Model	
	Efficiency Score	β	Efficiency Score	β

[i]When the first type of direction vector is used, that is, "a) =(-x0, y0, -b0)", the two parameters "total weight of good outputs (wg)" and "total weight of bad outputs (wb)" have no effect (setting them or not does not affect the results), so the software displays them as unavailable. When another direction vector type is selected, these two parameters can be set.

[ii]The author recommends setting undesirable outputs as strongly disposable, for the reasons analysed above.

Guangxi	0.9984	0.0008	0.9984	0.0008
Guizhou	0.9902	0.0049	0.9902	0.0049
Hainan	**0.9323**	**0.0350**	**0.9885**	**0.0058**
Hebei	0.9543	0.0234	0.9543	0.0234
Henan	**0.9508**	**0.0252**	**1.0000**	**0.0000**

4.5.7 A Common Misunderstanding about the Results of DEA Models with Undesirable Outputs

The following is a common misunderstanding about the results produced by DEA models with undesirable outputs.

Model 1: a radial or directional distance function model that does not include undesirable outputs;

Model 2: Model 1 plus one or more undesirable output measures.

The analysis shows that the efficiency scores obtained from Model 2 are always greater than or equal to those obtained from Model 1 (the scores of Model 2 may remain unchanged, but they never fall). Many beginners find such a result difficult to understand. Once undesirable outputs are added, should the efficiency score not fall?

This misunderstanding stems from confusing the two concepts of "technical efficiency" and "productivity". If productivity is being measured, the productivity obtained when undesirable outputs are included is lower than that obtained when they are excluded. Technical efficiency, however, is not productivity: it is a relative efficiency (relative to the production frontier) and can be regarded as the ratio of the productivity of the evaluated DMU to the productivity of the production frontier, a meaning similar to that of a productivity index. High productivity does not necessarily imply a high productivity index, and vice versa. When undesirable outputs are included, the technical efficiency obtained is higher because the productivity of the evaluated DMU falls by a smaller proportion than the productivity of the technological frontier, so that the ratio of the two (technical efficiency) rises.

DEA methods cannot measure productivity (or total factor productivity, TFP) itself; they can

measure only a TFP index (see Chapter 8 for details). Cumulative TFP indices can be computed through global or fixed benchmarking. A cumulative TFP index is still an index: accumulation changes only the reference period, from period t relative to period t-1 to period t relative to period t0 (the base period). Cumulating and directly computing the TFP index of period t relative to period t0 give the same result. Both the TFP index and the cumulative TFP index can only reflect changes in TFP and cannot yield the magnitude of TFP itself.

The computation process is now examined in order to see why the technical efficiency score rises once undesirable outputs are added.

It was noted earlier that, with the number of DMUs fixed, the more input and output measures a model contains, the higher the efficiency score obtained. If one input or output measure is added to an existing model (whether a CCR or a BCC model), the efficiency score of the new model is always greater than or equal to that of the original model. More generally, regardless of whether the added measure is an input or an output, and whether it is a desirable or an undesirable output, the efficiency score of the new model is always greater than or equal to that of the original model. The same conclusion applies to directional distance function models. This is a comparison between two DEA models.

The above conclusion can be proved simply.

Model 1:

$$\max \ \beta$$

$$s.t. \ X\lambda + \beta g_x \le x_k$$

$$Y\lambda - \beta g_y \ge y_k$$

$$g_x \ge 0, g_y \ge 0 \quad (4.50)$$

Model 2:

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda+\beta g_b \le b_k$$

$$g_x \ge 0, g_y \ge 0, g_b \ge 0 \quad (4.51)$$

Compared with Model 1, the only difference in Model 2 is that one or more constraints on the undesirable outputs are added $B\lambda-\beta g_b \ge b_k$. The objective function of the model is to maximize β, and the maximum β attainable by Model 2 after one or more constraints are added certainly does not exceed the maximum β obtained in Model 1. The larger β is, the smaller the efficiency score, so the efficiency score obtained from Model 2 is certainly greater than or equal to that obtained from Model 1.

The idea that technical efficiency should fall once undesirable outputs are added really applies to one particular DMU. The precise statement should be: in a model that includes undesirable outputs, if the undesirable output of a given DMU increases (with the other DMUs held constant), the efficiency score of that DMU falls (the actual result may also remain unchanged, but it certainly does not rise). The above conclusion can be verified in a model with undesirable outputs by changing the value of the undesirable output of a single DMU.

MaxDEA offers a simpler approach that does not require analysing the DMUs one at a time. Taking Table 4-23 as an example, the data comprise two inputs (X1, X2), one desirable output (Y) and one undesirable output (B). Group 0 is the original data, group 1 has the undesirable output reduced and group 2 has the undesirable output increased. The example uses a non-oriented VRS directional distance function model, with the direction vector type "a) = (-x0, y0, -b0)".[i] On the "Advanced Models" interface of MaxDEA, select "Undesirable Output" and set B as an undesirable output; select "Group Benchmarking" and "h) Fixed Benchmarking", and choose group 0 as the benchmark group.[ii]

[i] is equivalent to radial distance.

[ii] "Fixed benchmarking" is a group benchmarking method newly added in version 8.12. Users of versions

Table 4-24 shows the changes in the efficiency scores after the undesirable output is reduced and increased respectively. For the example data, the efficiency score rises after the undesirable output is reduced and falls or remains unchanged after it is increased.

Table 4-23 Sample Data with an Undesirable Output

Cluster	DMU	X1	X2	Y	B
0	A	14	18	19	15
0	B	14	11	12	12
0	C	17	18	11	14
0	D	14	15	13	19
0	E	20	11	19	15
0	F	19	16	11	19
0	G	19	10	10	20
0	H	18	15	11	16
0	I	19	17	19	19
0	J	20	14	12	14
1	A-1	14	18	19	10
1	B-1	14	11	12	7
1	C-1	17	18	11	9
1	D-1	14	15	13	14
1	E-1	20	11	19	10
1	F-1	19	16	11	14
1	G-1	19	10	10	15
1	H-1	18	15	11	11
1	I-1	19	17	19	14
1	J-1	20	14	12	9
2	A-2	14	18	19	20
2	B-2	14	11	12	17
2	C-2	17	18	11	19
2	D-2	14	15	13	24
2	E-2	20	11	19	20
2	F-2	19	16	11	24
2	G-2	19	10	10	25
2	H-2	18	15	11	21
2	I-2	19	17	19	24
2	J-2	20	14	12	19

8.0-8.11 can click "Check for Updates" in the software to update free of charge to the latest 8.x version. Users of version 7.x or earlier can obtain the same results by selecting "Custom Benchmarking" on the "Advanced Models" screen and setting the DMUs of group 0 as the reference set.

Table 4-24 Changes in the Efficiency Score after Reducing and Increasing the Undesirable Output

DMU	Original Efficiency Score	After Reduction of the Undesirable Output		Increase in the Undesirable Output	
		Efficiency Score of Group 1	Change in Efficiency Score	Efficiency Score of Group 2	Change in Efficiency Score
A	1.000000	1.760870	0.760870	1.000000	0
B	1.000000	6.000000	5.000000	1.000000	0
C	0.770270	2.000000	1.229730	0.700000	-0.070270
D	1.000000	1.000000	0.000000	1.000000	0
E	1.000000	1.760870	0.760870	1.000000	0
F	0.612022	0.770270	0.158248	0.612022	0
G	1.000000	1.111111	0.111111	1.000000	0
H	0.658824	1.200000	0.541176	0.658824	0
I	1.000000	1.094595	0.094595	1.000000	0
J	0.810811	2.000000	1.189189	0.662921	-0.147889

4.6 Weighted Additive Distance (Including SBM Directional Distance Function and Non-radial Directional Distance Function)

All models in this section are presented under VRS.

The models introduced in this section all belong to the same distance type, weighted additive distance. They are additive models that adopt different weights; RAM and BAM can be regarded as special cases of the directional SBM (the SBM directional distance function) and of the non-radial directional distance function, that is, the RAM and BAM models can be regarded as a directional SBM or a non-radial directional distance function that adopts particular pseudo direction vector values (discussed in detail below).

Apart from the RAM and BAM models, whose results are reported as an "efficiency score", all other types of weighted additive model report an "inefficiency score", which ranges from 0 to infinity; for super-efficiency models, the inefficiency score ranges over the real line (from minus infinity to plus infinity). For the same data, the actual range of the inefficiency score obtained depends on the weights specified. The smaller the weight values, the smaller the inefficiency score. An inefficiency score cannot be used to compute the Malmquist index; only the Luenberger index can be computed (see Chapter 8 for details). If the Malmquist index is required, choose the RAM or BAM model.

4.6.1 Unweighted Additive Distance (Simple Additive Distance)

The objective function of the unweighted additive model is the direct sum of the slack variables of the input and output measures, and its expression is

$$\begin{aligned}
&\max \ \sum_{i=1}^{m} s^- + \sum_{r=1}^{q} s^+ \\
&s.t. \ X\lambda + s^- = x_k \\
&\quad Y\lambda - s^+ = y_k \\
&\quad e\lambda = 1 \\
&\quad \lambda, s^-, s^+ \geq 0 \quad (4.52)
\end{aligned}$$

The above expression contains a logical error. The slack variables of the inputs and outputs

carry units, and adding slacks with different units directly is logically unsound.

Revising the model(4.52) into the simple-weight additive model of the next section (with weight values equal to 1) eliminates the above logical error.

When the model includes undesirable outputs,

$$\max \sum_{i=1}^{m} s^{-} + \sum_{r=1}^{q_1} s^{+} + \sum_{t=1}^{q_2} s^{b-}$$

$$s.t.\ X\lambda + s^{-} = x_k$$

$$Y\lambda - s^{+} = y_k$$

$$\mathrm{B}\lambda + s^{b-} = b_k$$

$$e\lambda = 1$$

$$\lambda, s^{-}, s^{+} \geq 0 \quad (4.53)$$

In this type of model, treating the undesirable outputs as inputs yields the same results.

4.6.2 Weighted Additive Distance

The weighted additive model builds on the additive model by assigning a different weight w to each slack variable in the objective function, so that the objective function is the weighted sum of the slack variables, and its expression is

$$\max \sum_{i=1}^{m} w_i^{-} s_i^{-} + \sum_{r=1}^{q} w_r^{+} s_r^{+}$$

$$s.t.\ X\lambda + s^{-} = x_k$$

$$Y\lambda - s^{+} = y_k$$

$$e\lambda = 1$$

$$\lambda, s^{-}, s^{+} \geq 0 \quad (4.54)$$

Note that in the above expression the slack weights w are dimensional, their dimension being the reciprocal of the dimension (unit of measurement) of the corresponding input or output. For example, if x1 is measured in kg, then w1 is measured in 1/kg.

When the numerical parts of all the weights w equal 1, the model reduces to the

"simple-weight additive model". Since the weights are dimensional, the simple-weight additive model does not suffer from the logical error of the unweighted additive model mentioned above.

Unit invariance means that the results of a model remain unchanged when the units of measurement of the input or output data are changed. The unrevised additive model, namely the model(4.52) does not satisfy unit invariance. Suppose that an input is originally measured in kilograms (kg); when the unit is changed to grams (g), the values of that input become 1000 times larger, and once the data units change, the analysis results of the model(4.52) change accordingly.

In the weighted additive model, because the weights incorporate the dimension, the model results remain unchanged when the data units change. Suppose that an input is originally measured in kilograms (kg) with a weight of 0.5/kg; when the unit is changed to grams (g), its weight becomes 0.5/kg = 0.5/(1000 g) = 0.0005/g, and the analysis results of the model(4.54) remain unchanged. When the data units change, both the numerical part and the dimensional part of the weight change accordingly: the numerical part changes from 0.5 to 0.0005 and the dimensional part from 1/kg to 1/g, so the actual weight remains unchanged.

Note that the weight set in the software is only the numerical part of the weight; if the units of measurement of the data change, the user should modify the numerical part of the weight accordingly. If the numerical part is not modified, the actual weight changes and the analysis results change with it.

If undesirable outputs are included in the weighted additive model, the expression is

$$\max \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q_1} w_r^+ s_r^+ + \sum_{t=1}^{q_2} w_t^{b-} s_t^{b-}$$

$$s.t.\ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$\mathrm{B}\lambda + s^{b-} = b_k$$

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.55)$$

It can be seen that in the weighted additive model with undesirable outputs, treating the

undesirable outputs as inputs yields the same results.

In MaxDEA, users can define the weights w themselves.

h) Custom (the same weights for all DMUs)

i) Custom (different weights for each DMU)

Option h requires the weights to be entered directly, whereas option i adds the corresponding weight variables to the original data.

When sub-option h is used, the weight value is entered directly for each input and output (the default weight value in the software is 1), and the weights set apply to all DMUs;

when sub-option i is used, a weight variable is selected for each input and output, so that different weights can be set for different DMUs. With sub-option i, the corresponding weight variables must be added to the data for each input and output before the data are imported. Note that the weight variables must be left as "Not defined" when the data are defined.

Besides custom weights, one of several common weight types can also be selected, including:

b) Standardized weights (Lovell & Pastor, 1995)(Lovell & Pastor, 1995)

the reciprocal of the sample standard deviation of the variable is used as the weight

c) Weight = 1/|x0|, 1/|y0|

the reciprocal of the absolute value of the evaluated DMU's own data is used as the weight

d) Weight = $1/|\bar{x}|$, $1/|\bar{y}|$

the reciprocal of the absolute value of the variable mean is used as the weight

e) Range adjusted (RAM) (W. W. Cooper, Park, & Pastor, 1999)(W. W. Cooper, Park, & Pastor, 1999)

f) Bounded adjusted (BAM) (William W. Cooper, Pastor, Borras, Aparicio, & Pastor, 2011)(William W. Cooper, Pastor, Borras, Aparicio, & Pastor, 2011)

g) Directional SBM (Fukuyama & Weber, 2009)(Fukuyama & Weber, 2009)

RAM, BAM and directional SBM are introduced separately below.

4.6.3 A Simple-Weight Variant: Slacks-based Measure on the Sum of Directional Distance Functions

The slacks-based measure on the sum of directional distance functions was proposed byR. Färe & Grosskopf (2010) as a variant of the simple additive model; in essence it is a weighted additive model with weight values equal to 1. The model is expressed as

$$\max \ \beta_1^x + ... + \beta_m^x + \beta_1^y + ... + \beta_q^y$$

$$s.t. \ \sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x e_i^x \le x_{i0}, i = 1,2,3,...,m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y e_r^y \ge y_{r0}, r = 1,2,3...,q$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda, \beta^x, \beta^y \ge 0 \quad (4.56)$$

e consists of two parts, a numerical part and a dimensional part; its numerical parts all equal 1, and its dimension is the same as that of the corresponding input or output. β is dimensionless.

Note that if the inequality sign in the model(4.56) is changed to an equality sign, the model results remain unchanged[i], that is,

[i]The constraint on the m-th input is taken as an example to show that in this model the inequality and the equality are equivalent.

$\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x \le x_{m0}$ is equivalent to $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x + \sigma_m = x_{m0}$, $\sigma_m \ge 0$. Since the objective function maximizes the sum of the β terms, suppose that in some solution of the model there is $\sigma_m^* > 0$, then it is always possible, while keeping $\beta_i \ (i \ne m)$ the optimal solution $\beta_i^* \ (i \ne m)$ unchanged, to obtain a better solution $\beta_m^{*2} = \beta_m^* + \sigma_m^*, \sigma_m^{*2} = 0$. Therefore the optimal solution of the model necessarily has $\sigma_m^* = 0$. It follows that $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x \le x_{m0}$ is equivalent

$$\max \ \beta_1^x + \ldots + \beta_m^x + \beta_1^y + \ldots + \beta_q^y$$

$$s.t. \ \sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x e_i^x = x_{ik}, i = 1,2,3,\ldots,m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y e_r^y = y_{rk}, r = 1,2,3\ldots,q$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda, \beta^x, \beta^y \geq 0 \quad (4.57)$$

Setting s = βe and w = 1/e, the model(4.57) can be transformed into

$$\max \ \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q} w_r^+ s_r^+$$

$$s.t. \ X\lambda + s^- = x_k$$

$$Y\lambda - s^+ = y_k$$

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.58)$$

The transformed model and the weighted additive model(4.54) are exactly the same.

Therefore, to implement the slacks-based measure on the sum of directional distance functions in MaxDEA, it suffices to select the first sub-option under the weighted additive model, namely "a) Simple weights: weight = (1, 1, 1,...)".

4.6.4 The RAM Model of Weighted Additive Distance

The RAM (range adjusted measure) model was proposed byW. W. Cooper et al. (1999) as a weighted additive model; a notable difference from the other weighted additive models is that the

---

to $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x = x_{m0}$.

RAM model defines an efficiency score, with values between 0 and 1[i].

$$\min \ 1-\frac{1}{m+q}\left(\sum_{i=1}^{m}\frac{s_i^-}{R_i^-}+\sum_{r=1}^{q}\frac{s_r^+}{R_r^+}\right)$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+\geq 0$$

$$R_i^-=\max(\mathrm{x}_i)-\min(\mathrm{x}_i)$$

$$R_r^+=\max(\mathrm{y}_r)-\min(\mathrm{y}_r) \quad (4.59)$$

When the model includes undesirable outputs,

$$\min \ 1-\frac{1}{m+q_1+q_2}\left(\sum_{i=1}^{m}\frac{s_i^-}{R_i^-}+\sum_{r=1}^{q_1}\frac{s_r^+}{R_r^+}+\sum_{t=1}^{q_2}\frac{s_t^{b-}}{R_t^{b-}}\right)$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+\geq 0$$

$$R_i^-=\max(\mathrm{x}_i)-\min(\mathrm{x}_i)$$

$$R_r^+=\max(\mathrm{y}_r)-\min(\mathrm{y}_r)$$

$$R_t^{b-}=\max(b_t)-\min(b_t) \quad (4.60)$$

[i]The RAM model is designed for use with the VRS model. In the CRS and NIRS models the maximum possible improvement in the input measures of the DMU k under evaluation may exceed Ri, that is, sik>Ri, so that the objective function may become negative. The RAM model is therefore not applicable to the CRS and NIRS models.

In the RAM model, treating the undesirable outputs as inputs yields the same results.

4.6.5 The BAM Model of Weighted Additive Distance

The BAM (bounded adjusted measure) model was proposed byWilliam W. Cooper et al. (2011) as a weighted additive model similar to the RAM model, whose efficiency score also lies between 0 and 1.[i]

$$\min \ 1-\frac{1}{m+q}\left(\sum_{i=1}^{m}\frac{s_i^-}{L_{ik}^-}+\sum_{r=1}^{q}\frac{s_r^+}{U_{rk}^+}\right)$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+\geq 0$$

$$L_{ik}^-=x_{ik}-\min(\mathrm{x}_i)$$

$$U_{rk}^+=\max(\mathrm{y}_r)-y_{rk} \quad (4.61)$$

If $L_{ik}^-=0$ then the corresponding s_i^-/L_{ik}^- term is dropped from the objective function; if $U_{rk}^+=0$ then the corresponding s_r^+/U_{rk}^+ term is dropped from the objective function.

$L_{ik}^-\leq R_i^-, U_{rk}^+\leq R_r^+$ so the efficiency score of the BAM model is less than or equal to that of the RAM model.

When the model includes undesirable outputs,

[i]The BAM model is designed for use with the VRS model. In the CRS and NIRS models the maximum possible improvement in the input measures of the DMU k under evaluation may exceed Lik, that is, sik>Lik, so that the objective function may become negative. The RAM model is therefore not applicable to the CRS and NIRS models.

$$\min \ 1-\frac{1}{m+q_1+q_2}\left(\sum_{i=1}^{m}\frac{s_i^-}{L_{ik}^-}+\sum_{r=1}^{q_1}\frac{s_r^+}{U_{rk}^+}+\sum_{t=1}^{q_2}\frac{s_t^{b-}}{L_{tk}^{b-}}\right)$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+\geq 0$$

$$L_{ik}^-=x_{ik}-\min(\mathrm{x}_i)$$

$$U_{rk}^+=\max(\mathrm{y}_r)-y_{rk}$$

$$L_{tk}^{b-}=b_{tk}-\min(\mathrm{b}_t) \quad (4.62)$$

In the BAM model, treating the undesirable outputs as inputs yields the same results.

4.6.6 The Directional SBM of Weighted Additive Distance (Directional Slacks-based Measure)

The directional SBM (also called the SBM directional distance function) was proposed byFukuyama & Weber (2009) as a model whose expression is

$$\max \ \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^-/g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+/g_r^y)}{2}$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+\geq 0 \quad (4.63)$$

(g^x, g^y) is the pseudo direction vector. Note that g^x, g^y are all positive and dimensional,

with the same dimension as the corresponding input or output, so the directional SBM also satisfies the unit invariance discussed above.

If its objective function is rewritten in a different form, it becomes clear that the directional SBM is essentially a weighted additive model.

$$\begin{aligned}\max \ & \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+ / g_r^y)}{2} \\ =& \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+ / g_r^y)}{2} \\ =& \sum_{i=1}^{m}\frac{1}{2mg_i^x}s_i^- + \sum_{r=1}^{q}\frac{1}{2qg_r^y}s_r^+\end{aligned} \quad (4.64)$$

It can be seen that the directional SBM model is a weighted additive model whose weights are $w^x = \frac{1}{2mg^x}$ and $w^y = \frac{1}{2qg^y}$ respectively.

MaxDEA provides the directional SBM directly as a sub-option under the weighted additive model, namely the item "g) Directional SBM(Fukuyama & Weber, 2009)". With this sub-option, users can construct the directional SBM model more conveniently, since only the pseudo direction vector values (gx and gy) have to be set. MaxDEA provides several types of pseudo direction vector for direct selection, and users can also customize the pseudo direction vector.

The fourth item, "(d) Range (RDM, Portela, Thanassoulis, and Simpson 2004)", draws on the RDM directional distance function approach proposed byPortela et al. (2004) where the pseudo direction vector is the maximum possible improvement of the evaluated DMU (range of possible improvement), that is,

$$g_{xk} = x_k - \min(\mathrm{x})$$

$$g_{yk} = \max(\mathrm{y}) - y_k$$

For undesirable outputs, $g_{bk} = b_k - \min(b)$

if VRS is selected, the inefficiency score obtained with this type of pseudo direction vector lies between 0 and 1 and can be converted conveniently into an efficiency score, for example by "efficiency score = 1 - inefficiency score".

If custom pseudo direction vectors are used (options e and f), note that the pseudo direction vector values are positive whether the measure is an input or an output and whether it is a desirable or an undesirable output.

When the model includes undesirable outputs,

$$\max \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}(s_r^+ / g_r^y)+\sum_{t=1}^{q_2}(s_t^{b-} / g_t^b)\right)}{2}$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda, s^-, s^+ \geq 0 \quad (4.65)$$

In the objective function of the directional SBM model, the number of input measures m and the number of output measures q (or q1 + q2) are not combined as they are in the RAM and BAM models. Therefore, treating the undesirable outputs as inputs in the directional SBM model gives different results.

The directional SBM model yields an inefficiency score, so if a TFP index is to be computed for panel data, only the Luenberger index approach (addition and subtraction) can be used. How, then, can an efficiency score be obtained and the Malmquist index (multiplication and division) be used?

The relationship between the directional SBM model and the RAM and BAM models is examined below.

The directional SBM, RAM and BAM models all belong in essence to the weighted additive model, and each of the three can be regarded as a weighted additive model that adopts different weights; RAM and BAM convert the inefficiency score in the results into an efficiency score on the basis of the weighted additive model (conversion: efficiency score = 1 - inefficiency score).

The inefficiency score obtained from the directional SBM model

$$= \max \sum_{i=1}^{m} \frac{1}{2mg_i^x} s_i^- + \sum_{r=1}^{q} \frac{1}{2qg_r^y} s_r^+$$

the efficiency score obtained from the RAM model is 1 - inefficiency score

$$= \min 1 - \left(\sum_{i=1}^{m} \frac{s_i^-}{(m+q)R_i^-} + \sum_{r=1}^{q} \frac{s_r^+}{(m+q)R_r^+} \right)$$

$$= \min 1 - \left(\sum_{i=1}^{m} \frac{s_i^-}{2\,\mathrm{m}(m+q)R_i^- / (2m)} + \sum_{r=1}^{q} \frac{s_r^+}{2\,\mathrm{q}(m+q)R_r^+ / (2q)} \right)$$

Relationship between the RAM model and the directional SBM model: the RAM model can be regarded as a special case of the directional SBM model, that is, the RAM model corresponds to a directional SBM model with pseudo direction vectors $g_i^x = \frac{(m+q)R_i^-}{2m}, g_r^y = \frac{(m+q)R_r^+}{2q}$ that is converted by the "efficiency score = 1 - inefficiency score" method.

Relationship between the BAM model and the directional SBM model: the BAM model can be regarded as a special case of the directional SBM model, that is, the BAM model corresponds to a directional SBM model with pseudo direction vectors $g_i^x = \frac{(m+q)L_{ik}^-}{2m}, g_r^y = \frac{(m+q)U_{rk}^+}{2q}$ that is converted by the "efficiency score = 1 - inefficiency score" method.

If the name "directional SBM model" is to be retained while an "efficiency score" is required (so that the Malmquist index can be computed), the RAM or BAM model can be chosen, since RAM and BAM can be regarded as "directional SBM models" whose pseudo direction vectors are set to the specific values above and whose inefficiency scores are converted into efficiency scores.

Alternatively, the inefficiency scores output by the directional SBM model can be converted into efficiency scores afterwards by a user-defined method, and the Malmquist index recomputed manually (the conversion method must ensure that the resulting efficiency scores are positive, otherwise the Malmquist index cannot be computed).

4.6.7 The Non-radial Directional Distance Function of Weighted Additive Distance

This is a non-radial directional distance function widely used in the literature; see(Barros, Managi, & Matousek, 2012; Zhang, Zhou, & Choi, 2013; Mirsalehy, Abu Bakar, Lee, Jaafar, & Heydar, 2014; Yao, Zhou, Zhang, & Li, 2015; Lee & Choi, 2018).

Its expression is

$$\max \sum_{i=1}^{m}\omega_i^x\beta_i^x+\sum_{r=1}^{q_1}\omega_r^y\beta_r^y+\sum_{t=1}^{q_2}\omega_t^b\beta_t^b$$

$$s.t.\ \sum_{j=1}^{n}x_{ij}\lambda_j+\beta_i^x g_i^x\le x_{ik},i=1,2,3,...,m$$

$$\sum_{j=1}^{n}y_{rj}\lambda_j-\beta_r^y g_r^y\ge y_{rk},r=1,2,3...,q_1$$

$$\sum_{j=1}^{n}b_{tj}\lambda_j+\beta_t^b g_t^b=b_{tk},t=1,2,3...,q_2$$

$$\sum_{j=1}^{n}\lambda_j=1$$

$$\lambda,\beta^x,\beta^y,\beta^b\ge 0 \quad (4.66)$$

$(-g^x,g^y,-g^b)$ is the pseudo direction vector.

The so-called direction vector in the model is referred to here as a pseudo direction vector because the model contains no true direction vector, only something that resembles one. The above model cannot be expressed compactly in matrix and vector form as a genuine directional distance function model can.

Note that if the inequality sign in the model(4.66) is changed to an equality sign, the model results remain unchanged[i], that is,

[i]The constraint on the m-th input is taken as an example to show that in this model the inequality and the equality are equivalent.

$\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x\le x_{m0}$ is equivalent to $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x+\sigma_m=x_{m0}$, $\sigma_m\ge 0$.

Since the objective function maximizes the weighted sum of the β terms, suppose that in some

$$\max \sum_{i=1}^{m}\omega_i^x\beta_i^x+\sum_{r=1}^{q_1}\omega_r^y\beta_r^y+\sum_{t=1}^{q_2}\omega_t^b\beta_t^b$$

$$s.t.\ \sum_{j=1}^{n}x_{ij}\lambda_j+\beta_i^x g_i^x=x_{ik},i=1,2,3,...,m$$

$$\sum_{j=1}^{n}y_{rj}\lambda_j-\beta_r^y g_r^y=y_{rk},r=1,2,3...,q_1$$

$$\sum_{j=1}^{n}b_{ij}\lambda_j+\beta_t^b g_t^b=b_{tk},t=1,2,3...,q_2$$

$$\sum_{j=1}^{n}\lambda_j=1$$

$$\lambda,\beta^x,\beta^y,\beta^b\geq 0 \quad (4.67)$$

Setting s = βg and w = ω/g, the model(4.67) can be transformed into

$$\max \sum_{i=1}^{m}w_i^-s_i^-+\sum_{r=1}^{q_1}w_r^+s_r^++\sum_{t=1}^{q_2}w_t^{b-}s_t^{b-}$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda,s^-,s^+\geq 0 \quad (4.68)$$

The transformed model and the weighted additive model that includes undesirable outputs(4.55) are exactly the same.

In the non-radial directional distance function model, treating the undesirable outputs as

---

solution of the model there is $\sigma_m^*>0$, then it is always possible, while keeping $\beta_i \quad (i\neq m)$ the optimal solution $\beta_i^* \quad (i\neq m)$ unchanged, to obtain a better solution $\beta_m^{*2}=\beta_m^*+\sigma_m^*,\sigma_m^{*2}=0$. Therefore the optimal solution of the model necessarily has $\sigma_m^*=0$. It follows that $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x\leq x_{m0}$ is equivalent to $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x=x_{m0}$.

inputs yields the same results.

It can be seen that the above non-radial directional distance function model is a weighted additive model with weights w = ω/g.

Note that the pseudo direction vector of the non-radial directional distance function is $(-g^x, g^y, -g^b)$. g^x, g^y, g^b are all positive, so the resulting weights (w^x, w^y, w^b) are all positive.

The above transformation requires the pseudo direction vector values g > 0. When the pseudo direction vector value g of an input or output has to be set to 0, the corresponding weight of that input or output in the objective function is w = 0, and the corresponding constraints are as follows:

For an input, $\sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x g_i^x \le x_{ik}$ which gives $\sum_{j=1}^{n} x_{ij}\lambda_j \le x_{ik}$ that is, $X\lambda + s^- = x_k$;

this is the same as the input constraint term in the model(4.68).

For an output, $\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y g_r^y \ge y_{rk}$ which gives $\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rk}$ that is, $Y\lambda - s^+ = y_k$;

this is the same as the output constraint term(4.68) in the model.

For an undesirable output, $\sum_{j=1}^{n} b_{tj}\lambda_j + \beta_t^b g_t^b = b_{tk}$ which gives $\sum_{j=1}^{n} b_{tj}\lambda_j = b_{tk}$ which is equivalent to $B\lambda + s^{b-} = b_k,\ s^{b-} = 0$; the difference from the undesirable output constraint term(4.68) in the model is that an additional restriction $s^{b-} = 0$ is imposed.[i]

In MaxDEA, the above non-radial directional distance function can be implemented simply by selecting the "Custom" sub-options (sub-option h or i) under the weighted additive model.

h) Custom (the same weights for all DMUs)

i) Custom (different weights for each DMU)

First the weights are computed with the formulas above, that is, $w^x = \omega^x / g^x\ w^y = \omega^y / g^y\ w^b = \omega^b / g^b$.

and then the weights are set. Option h requires the computed weights to be entered directly,

[i] s = 0; the corresponding setting in MaxDEA is to select "inputs and outputs that cannot be freely controlled" among the advanced model options and to set the "slack freedom" of this undesirable output to 0.

whereas option i uses the computed weights as a weight variable in the data.

When sub-option h is used, the weight value is entered directly for each input and output (the default weight value in the software is 1), and the weights set apply to all DMUs;

when sub-option i is used, a weight variable is selected for each input and output, so that different weights can be set for different DMUs. With sub-option i, the corresponding weight variables must be added to the data for each input and output before the data are imported. Note that the weight variables must be left as "Not defined" when the data are defined.

The output of MaxDEA reports the Inefficiency Score and the slack movement of the inputs and outputs. If the β values of the individual variables are required, they can be computed with the following formula:

$$\beta_i = s_i^* / g_i \quad (4.69)$$

si* is the absolute value of the slack movement of each input and output in the results, and gi is the direction vector value of each input and output defined by the user. In the formula, si* ≥ 0 and gi > 0.

Taking Table 4-25 as an example, with (-x0, y0, -b0) as the pseudo direction vector, the weights (ω) of the two inputs (X1 and X2), the desirable output (Y) and the undesirable output (B) are 0.25, 0.25, 0.3 and 0.2 respectively.

Table 4-25 Sample Data for the Non-radial Directional Distance Function

DMU	X1	X2	Y	B	w_x1	w_x2	w_y	w_b
A	14	18	19	15	0.017857	0.013889	0.015789	0.013333
B	14	11	12	12	0.017857	0.022727	0.025000	0.016667
C	17	18	11	14	0.014706	0.013889	0.027273	0.014286
D	14	15	13	19	0.017857	0.016667	0.023077	0.010526
E	20	11	19	15	0.012500	0.022727	0.015789	0.013333
F	19	16	11	19	0.013158	0.015625	0.027273	0.010526
G	19	10	10	20	0.013158	0.025000	0.030000	0.010000
H	18	15	11	16	0.013889	0.016667	0.027273	0.012500
I	19	17	19	19	0.013158	0.014706	0.015789	0.010526
J	20	14	12	14	0.012500	0.017857	0.025000	0.014286

The final weights (w) of the measures are $w^x = \omega^x / g^x, w^y = \omega^y / g^y\ w^b = \omega^b / g^b$ that is, $w^{x1} = 0.25 / x_1\ w^{x2} = 0.25 / x_2$, $w^y = 0.3 / y\ w^b = 0.2 / b$.

Add the four calculated weights to the data as variables. Leave them as "Not defined" when defining the data. For the weight type, select "i) Custom (different weights for each DMU)" and set the weight variable for each measure; select VRS for returns to scale; on the "Advanced Models" interface, set B as an undesirable output.

Table 4-26 shows the analysis results and the Beta values calculated with the formula(4.69). Readers can use the β values and the weights ω (0.25, 0.25, 0.3, 0.2) to calculate the inefficiency score, that is, inefficiency = $\sum_{i=1}^{m}\omega_i^x\beta_i^x + \sum_{r=1}^{q_1}\omega_r^y\beta_r^y + \sum_{t=1}^{q_2}\omega_t^b\beta_t^b$ which can be checked against the inefficiency score reported by the software.

Table 4-26 Analysis Results of the Non-radial Directional Distance Function

DM	Inefficiency	Slack improvement[i]				Beta value			
U	score	X1	X2	Y	B	X1	X2	Y	B
A	0	0	0	0	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0
C	0.2356	0	-5.8333	5.6667	0	0	0.3241	0.5152	0
D	0.1249	0	0	3.0000	-5.2857	0	0	0.2308	0.2782
E	0	0	0	0	0	0	0	0	0
F	0.3202	0	-3.8333	8.0000	-4.0000	0	0.2396	0.7273	0.2105
G	0	0	0	0	0	0	0	0	0
H	0.2585	0	-1.6667	8.0000	-1.0000	0	0.1111	0.7273	0.0625
I	0.1132	0	-4.8333	0	-4.0000	0	0.2843	0	0.2105
J	0.1952	-2	-3.0000	4.6667	0	0.1000	0.2143	0.3889	0

The non-radial directional distance function yields inefficiency scores. If the TFP index for

[i]In the results of MaxDEA the sign of a slack improvement indicates the direction of the improvement, a positive number indicating an increase and a negative number a decrease.

panel data is to be computed, only the Luenberger index approach (additive and subtractive operations) can be used. How, then, can efficiency scores be obtained so that the Malmquist index (multiplicative and divisive operations) can be applied? The method is the same as that presented in the previous section.

The relationship of the non-radial directional distance function model to the RAM model and the BAM model is examined below.

The non-radial directional distance function model, the RAM model and the BAM model all belong essentially to the class of weighted additive models; each of the three can be regarded as a weighted additive model with a different set of weights. RAM and BAM build on the weighted additive model by converting the resulting inefficiency scores into efficiency scores (conversion: efficiency score = 1 - inefficiency score).

The inefficiency score obtained from the non-radial directional distance function model

$$= \max \sum_{i=1}^{m} \frac{\omega_i^-}{g_i^-} s_i^- + \sum_{r=1}^{q_1} \frac{\omega_r^+}{g_r^+} s_r^+$$

the efficiency score obtained from the RAM model equals 1 - inefficiency score

$$= \min \ 1 - \left(\sum_{i=1}^{m} \frac{s_i^-}{(m+q)R_i^-} + \sum_{r=1}^{q} \frac{s_r^+}{(m+q)R_r^+} \right)$$

$$= \min \ 1 - \left(\sum_{i=1}^{m} \frac{\omega_i^-}{(m+q)R_i^- / \omega_i^-} s_i^- + \sum_{r=1}^{q} \frac{\omega_r^+}{(m+q)R_r^+ / \omega_r^+} s_r^+ \right)$$

Relationship between the RAM model and the non-radial directional distance function model: the RAM model can be regarded as a special case of the non-radial directional distance function, that is, the RAM model is equivalent to a non-radial directional distance function model with pseudo-direction vector $g_i^x = \frac{(m+q)R_i^-}{\omega_i^-}, g_r^y = \frac{(m+q)R_r^+}{\omega_r^+}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

Relationship between the BAM model and the non-radial directional distance function model: the BAM model can be regarded as a special case of the non-radial directional distance function, that is, the BAM model is equivalent to a non-radial directional distance function model with

pseudo-direction vector $g_i^x = \frac{(m+q)L_{ik}^-}{\omega_i^-}, g_r^y = \frac{(m+q)U_{rk}^+}{\omega_r^+}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

If the name "non-radial directional distance function" is to be retained while efficiency scores (from which the Malmquist index can be computed) are also required, the RAM model or the BAM model may be chosen, since RAM and BAM can be regarded as non-radial directional distance functions in which the pseudo-direction vector is set to the specific values given above and the inefficiency scores are converted into efficiency scores.

Alternatively, a user-defined method can be applied afterwards to convert the inefficiency scores of the non-radial directional distance function output by the software into efficiency scores, and the Malmquist index can then be recalculated manually (the conversion method must ensure that the resulting efficiency scores are positive; otherwise the Malmquist index cannot be computed).

4.6.8 Relationships among the Various Weighted Additive Distances

All types of weighted additive distance belong essentially to the same class of distance.

1) Relationship between the Simple Additive Distance and the Directional SBM Model

The simple additive distance can be regarded as a special case of the directional SBM model, that is, it is equivalent to a directional SBM model with pseudo-direction vector $g_i^x = \frac{1}{2m}, g_r^y = \frac{1}{2q}$.

2) Relationship between the Simple Additive Distance and the Non-radial Directional Distance Function Model

The simple additive distance can be regarded as a special case of the non-radial directional distance function, that is, it is equivalent to a non-radial directional distance function model with pseudo-direction vector $g_i^x = \omega_i^-, g_r^y = \omega_r^+$.

3) Relationship between the RAM Model and the Directional SBM Model

The RAM model can be regarded as a special case of the directional SBM model, that is, the RAM model is equivalent to a directional SBM model with pseudo-direction vector

$g_i^x = \frac{(m+q)R_i^-}{2m}, g_r^y = \frac{(m+q)R_r^+}{2q}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

4) Relationship between the RAM Model and the Non-radial Directional Distance Function Model

The RAM model can be regarded as a special case of the non-radial directional distance function, that is, the RAM model is equivalent to a non-radial directional distance function model with pseudo-direction vector $g_i^x = \frac{(m+q)R_i^-}{\omega_i^-}$, $g_r^y = \frac{(m+q)R_r^+}{\omega_r^+}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

5) Relationship between the BAM Model and the Directional SBM Model

The BAM model can be regarded as a special case of the directional SBM model, that is, the BAM model is equivalent to a directional SBM model with pseudo-direction vector $g_i^x = \frac{(m+q)L_{ik}^-}{2m}, g_r^y = \frac{(m+q)U_{rk}^+}{2q}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

6) Relationship between the BAM Model and the Non-radial Directional Distance Function Model

The BAM model can be regarded as a special case of the non-radial directional distance function, that is, the BAM model is equivalent to a non-radial directional distance function model with pseudo-direction vector $g_i^x = \frac{(m+q)L_{ik}^-}{\omega_i^-}$, $g_r^y = \frac{(m+q)U_{rk}^+}{\omega_r^+}$ and converted using the "efficiency score = 1 - inefficiency score" rule.

4.7 The EBM Hybrid Distance

The EBM hybrid distance function was proposed byK. Tone & Tsutsui (2010) as a hybrid model that contains both radial and SBM distance functions; because the model uses an ε parameter, Tone called it the epsilon-based measure (EBM).

4.7.1 The EBM Model

The input-oriented CRS model is taken as an example below to explain how the EBM model is computed. The linear programming formulation of the input-oriented CRS EBM model is expressed as

$$\min \ \theta-\varepsilon\frac{1}{\sum_{i=1}^{\mathrm{m}} w_i^-}\sum_{i=1}^{m}\frac{w_i^- s_i^-}{x_{\mathrm{k}}}$$

$$s.t. \ X\lambda-\theta x_k+s^-=0$$

$$Y\lambda\geq y_{\mathrm{k}}$$

$$\lambda\geq 0, s^-\geq 0 \quad (4.70)$$

The efficiency score of the DMU under evaluation is the optimal solution of the objective function, that is,

$$\theta^*-\varepsilon\sum_{i=1}^{m}\frac{w_i^- s_i^{-*}}{x_{\mathrm{k}}}.$$

The model contains m+1 parameters: ε and wi- (i=1,2,...,m). wi- represents the relative importance of each input measure, and ε is a key parameter with a range of [0, 1] that represents the importance of the non-radial part in the calculation of the efficiency score: a value of 0 makes the model equivalent to a radial model, and a value of 1 makes it equivalent to the SBM model.

These parameter values must be determined before the EBM model is built.K. Tone & Tsutsui (2010) argued that these parameters should be determined by the data themselves and proposed a method for calculating them. Taking a model with two inputs (x1 and x2) as an example, the method is as follows:

First, the projected values of the input measures are obtained from the SBM model or

another model[i].

Here the two inputs are X1 and X2, and their projected values are denoted P1 and P2. The association between the projected values of the two measures represents the quantitative ratio of the two inputs under the production technology (production frontier) concerned. By analysing this quantitative relationship, the substitutability of the two inputs in the production process can be inferred: if the quantities of the two inputs are highly positively and linearly correlated, their substitutability is poor and production must be carried out at a relatively fixed ratio, so that production efficiency is measured mainly radially and ε should take a small value, possibly even 0; if the quantities of the two inputs are highly negatively and linearly correlated, their substitutability is strong and production need not follow a fixed ratio, so that production efficiency is measured mainly by non-radial models and ε should take a large value, possibly even 1.

An association matrix of the projected values of the input measures is then constructed, whose elements are the affinity indices between the projected values of each pair of input measures.

With two inputs, the matrix is expressed as:

Table 4-27 Affinity index matrix

	x1	x2
x1	S(P1,P1)	S(P1,P2)
x2	S(P2,P1)	S(P2,P2)

S( ) is the function that computes the affinity index between the projected values P1 and P2 of two input measures.K. Tone & Tsutsui (2010) argued that the affinity index should satisfy the following rules:

(1) Identical: S(a,a) = 1, that is, the affinity index of a measure with itself equals 1;

(2) Symmetric: S(a,b) = S(b,a), that is, the index is independent of the order of the two measures;

(3) Units-invariant: S(ta,b) = S(a,b) (t > 0), that is, the unit of measurement of the data has no effect on the index value;

[i]In MaxDEA the projected values are calculated with the SBM model.

(4) Range: 0≤S(a,b) ≤ 1, that is, the index lies between 0 and 1.

Following this line of reasoning, the first candidate that comes to mind is the linear correlation coefficient, namely the Pearson correlation coefficient.K. Tone & Tsutsui (2010) argued that the Pearson correlation coefficient fails to satisfy the above conditions.

K. Tone & Tsutsui (2010)'s method is first to compute the dispersion index between every pair of measures with a dispersion index function, and then to compute the affinity index from the dispersion indices:

S(a,b)=1-2D(a,b).

The dispersion index is computed as

$$D(a,b)=\begin{cases}\dfrac{\sum_{j=1}^{n}|c_j-\overline{c}|}{n(c_{\max}-c_{\min})} & (if\ c_{\max}>c_{\min}), \quad (4.71)\\ 0 & (if\ c_{\max}=c_{\min})\end{cases}$$

$c_j=\ln\frac{b_j}{a_j}$, $\overline{c}=\frac{1}{n}\sum_{j=1}^{n}\ln\frac{b_j}{a_j}$, $c_{\max}=\max(c_j)$, and $c_{\min}=\min(c_j)$.

Finally, the parameters required in the EBM are computed from the affinity index matrix thus constructed.

$$\varepsilon=\frac{m-\max(\rho)}{m-1}$$

$$w_i=\frac{v_i}{\sum_{i=1}^{m}v_i}$$

ρ is the largest eigenvalue of the affinity index matrix and v is its corresponding eigenvector.

The programming formulation of the output-oriented EBM model is expressed as

$$\min \frac{1}{\phi+\varepsilon^{+}\frac{1}{\sum_{r=1}^{q} w_r^{+}}\sum_{r=1}^{q}\frac{w_r^{+}s_r^{+}}{y_{\mathrm{k}}}}$$

$$s.t.\ X\lambda \le x_k$$

$$Y\lambda-\phi y_{\mathrm{k}}-s^{+}=0$$

$$\lambda \ge 0, s^{-} \ge 0 \quad (4.72)$$

The programming formulation of the non-oriented EBM model is expressed as

$$\min \frac{\theta-\varepsilon^{-}\frac{1}{\sum_{i=1}^{\mathrm{m}} w_i^{-}}\sum_{i=1}^{m}\frac{w_i^{-}s_i^{-}}{x_{\mathrm{k}}}}{\phi+\varepsilon^{+}\frac{1}{\sum_{r=1}^{q} w_r^{+}}\sum_{r=1}^{q}\frac{w_r^{+}s_r^{+}}{y_{\mathrm{k}}}}$$

$$s.t.\ X\lambda-\theta x_k+s^{-}=0$$

$$Y\lambda-\phi y_{\mathrm{k}}-s^{+}=0$$

$$\lambda \ge 0, s^{-} \ge 0 \quad (4.73)$$

4.7.2 Shortcomings of the EBM Model and Methods of Improvement

Problem 1: the projected values of the EBM model may involve logical errors.

In the input-oriented EBM model(4.70) the range of the variable θ is unrestricted, andK. Tone & Tsutsui (2010) did not discuss this either. When the range of the variable θ is unconstrained, the optimal solution θ* of the model may be greater than 1. The projected value of an input measure of an inefficient DMU may then be higher than its original value, that is, the improvement target may erroneously be to "increase the input".

Similarly, in the output-oriented EBM model(4.72) with the range of the variable φ unconstrained, the optimal solution φ* of the model may be less than 1. The projected value of an output measure of an inefficient DMU may then be lower than its original value, that is, the improvement target may erroneously be to "reduce the output".

In the non-oriented EBM model(4.73) the problem is more serious, and both types of error

may occur.

Here the Table 4-21 data are used to demonstrate this problem, with the VRS model adopted and the input-oriented, output-oriented and non-oriented EBM models solved separately, the EBM parameters being obtained with the affinity index method proposed byK. Tone & Tsutsui (2010) in that study. The erroneous results are shown in Table 4-28,Table 4-29,Table 4-30.

To avoid such errors, θ and φ must be constrained in the EBM model; taking the non-oriented EBM model as an example, the revised formulation is

$$\min \frac{\theta-\varepsilon^{-}\dfrac{1}{\sum_{i=1}^{\mathrm{m}} w_i^{-}}\sum_{i=1}^{m}\dfrac{w_i^{-}s_i^{-}}{x_{\mathrm{k}}}}{\phi+\varepsilon^{+}\dfrac{1}{\sum_{r=1}^{q} w_r^{+}}\sum_{r=1}^{q}\dfrac{w_r^{+}s_r^{+}}{y_{\mathrm{k}}}}$$

$$s.t.\ \ X\lambda-\theta x_k+s^{-}=0$$

$$Y\lambda-\phi y_{\mathrm{k}}-s^{+}=0$$

$$\lambda\geq 0, s^{-}\geq 0, \theta\leq 1, \phi\geq 1 \quad (4.74)$$

Relationship of the EBM model to the radial and SBM models:

1) If ε = 0, the EBM model is identical to the radial model.

2) If ε = 1, the EBM model is equivalent to the (weighted) SBM model; in this case, because the variables θ and/or φ are added, the model may have infinitely many optimal solutions, but all of them yield the same efficiency score and projected values.

Table 4-28 Erroneous results of the EBM model (input-oriented VRS)

Region	Efficiency score	θ	Number of beds Original value	Number of beds Projected value
Guangxi	0.9834	1.0001	95752	95760
Jiangxi	0.9979	1.0037	87184	87504
Tianjin	0.9855	1.0295	40787	41991

Table 4-29 Erroneous results of the EBM model (output-oriented VRS)

Region	Efficiency score	φ	Number of admissions Original value	Number of admissions Projected value
Anhui	0.8414	0.9929	4391516	4360170
Guangxi	0.9217	0.9749	3258438	3176512
Guizhou	0.8257	0.8953	2433370	2178564
Henan	0.8412	0.9580	7043628	6747705
Hubei	0.8907	0.9984	4805124	4797549
Hunan	0.8263	0.9408	5393968	5074549
Jiangsu	0.9520	0.9949	6433837	6401057
Jiangxi	0.8780	0.9097	3000964	2729925
Shandong	0.8318	0.9370	8354234	7827682
Sichuan	0.9196	0.9199	6489314	5969408
Xinjiang	0.8625	0.8795	3115413	2740144

Table 4-30 Erroneous results of the EBM model (non-oriented VRS)

Region	Efficiency score	θ	φ	Original number of beds	Projected number of beds	Number of health technical personnel Original value	Projected number of health technical personnel	Original number of patient visits	Projected number of patient visits	Number of admissions Original value	Number of admissions Projected value
Anhui	0.8407	**1.1771**	1.1718	140997	165965	132739	156245				
Gansu	0.8048	**2.0742**	2.4323	66661	138266	53127	110195				
Guangxi	0.9163	**2.2404**	2.1091	95752	214524	105773	236975				
Guizhou	0.8209	**1.9629**	1.9375	78368	153830	69320	136070				
Hainan	0.7464	**11.5154**	13.0687	21367	246050	26329	289388				
Hebei	0.7959	0.9346	**0.9649**							5675067	5475952
Henan	0.8112	0.7144	**0.7576**							7043628	5336206
Heilongjiang	0.6273	**1.3622**	1.8426	129449	176338	127358	173489				
Hubei	0.8903	**1.1378**	1.1231	152062	173022	147628	167977				
Hunan	0.8232	0.6101	**0.5290**							5393968	2853139
Jilin	0.6934	**1.6842**	2.1743	94636	159390	86278	145313				
Jiangsu	0.9095	0.6091	**0.6697**					166944454	111801185	6433837	4308682
Jiangxi	0.8685	**2.3167**	2.1414	87184	201978	93287	216116				
Inner Mongolia	0.6604	**2.5478**	3.3446	72871	185659	74177	188987				
Ningxia	0.7982	**8.0824**	9.2620	22037	178112	21830	176439				

Problem 2: the affinity index is computed by a wrong method, which cannot reflect the true degree of association and may even produce completely contrary results.

K. Tone & Tsutsui (2010) pointed out that the computation of the affinity index has the following feature: when the projected values of two measures are extremely radially distributed, the affinity index should equal 1, whereas in the extremely dispersed case it should equal 0 (see

Figure 4-11).K. Tone & Tsutsui (2010) gave two simple examples to illustrate this feature; in the extremely dispersed example (Example 2 in that paper), the computed affinity index is 0. That example, however, is too simple, involving only two efficient DMUs, and cannot make the point.

A simple example of extreme dispersion is given here, namely Table 4-31, to show that the affinity index computation method proposed byK. Tone & Tsutsui (2010) does not conform to the above requirement. As shown in Figure 4-11, the projected values of the two input measures are perfectly negatively and linearly correlated.[i]

Table 4-31 Sample data for computing the affinity index (1)

DMU	x1	x2	y
A	1	5	1
B	2	4	1
C	3	3	1
D	4	2	1
E	5	1	1
F	3	4	1
G	4	3	1

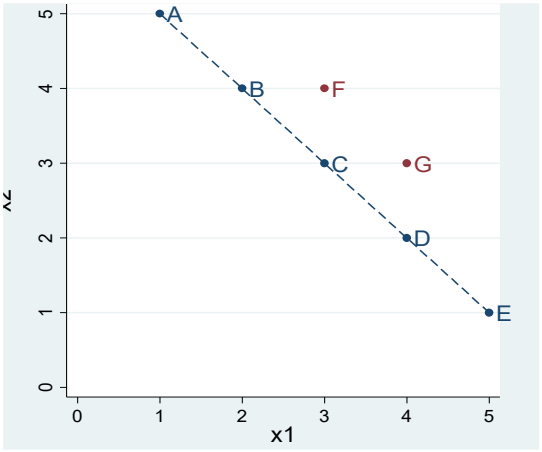


Figure 4-11 Example of dispersed data

Using the affinity index computation method described above, the affinity index obtained is S(P1,P2) = 0.428 and ε = 0.572. This result is inconsistent with theK. Tone & Tsutsui (2010) theory on which the affinity index is constructed.

Table 4-32 Dispersion index matrix for the sample data

	x1	x2
x1	0	0.286
x2	0.286	0

[i]The projected value of the inefficient DMU is one of the five efficient DMUs.

Table 4-33 Affinity index matrix for the sample data

	x1	x2
x1	1	0.428
x2	0.428	1

The example above shows that this method of computing the affinity index cannot reflect the true degree of association. Two further examples that yield completely wrong results are examined below.

Table 4-34 shows two almost perfectly associated variables, x1 and x2, as shown in

Figure 4-12, with an error of only 0.0001%. However, usingK. Tone & Tsutsui (2010) method, the dispersion index obtained is D(x1, x2) = 0.5 and the affinity index is S(x1, x2) = 0.

Table 4-35 shows two highly associated variables, x1 and x2, as shown in Figure 4-13. However, applying theK. Tone & Tsutsui (2010) method likewise yields the dispersion index D(x1, x2) = 0.5 and the affinity index S(x1, x2) = 0.

These two examples show that the method used byK. Tone & Tsutsui (2010) to compute the dispersion index and the affinity index is wrong: the dispersion index formula(4.71) reflects not the degree of dispersion of the two variables but the ratio of the mean dispersion ($\frac{\sum_{j=1}^{n}|c_j-\overline{c}|}{n}$) to the range of dispersion ($c_{\max}-c_{\min}$). The two examples above show that whenever the mean dispersion is half the range of dispersion, however small that range may be (as long as it is not 0), the resulting dispersion index D(x1, x2) always equals its maximum possible value of 0.5, and hence the affinity index is S(x1, x2) = 0. Whenever the mean dispersion approaches half the range of dispersion, the dispersion index D(x1, x2) approaches 0.5 and the affinity index S(x1, x2) approaches 0.

Table 4-34 Example of associated data (1)

x1	x2	x1/x2	c=ln(x1/x2)
1.000000	0.999999	1.000001	0.000001
2.000000	1.999998	1.000001	0.000001
3.000000	2.999997	1.000001	0.000001
4.000000	3.999996	1.000001	0.000001
5.000000	4.999995	1.000001	0.000001
6.000000	6.000006	0.999999	-0.000001
7.000000	7.000007	0.999999	-0.000001
8.000000	8.000008	0.999999	-0.000001
9.000000	9.000009	0.999999	-0.000001
10.000000	10.00001	0.999999	-0.000001

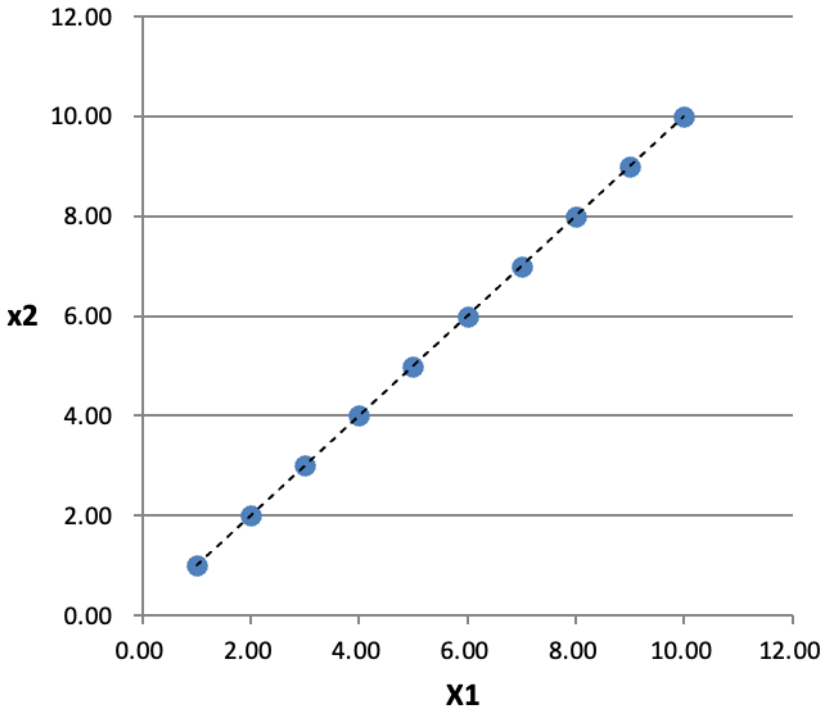


Figure 4-12 Example of associated data (1)

Table 4-35 Example of associated data (2)

x1	x2	x1/x2	c=ln(x1/x2)
1.000000	0.999999	1.000001	0.000001
2.000000	1.999998	1.000001	0.000001
3.000000	2.999997	1.000001	0.000001
4.000000	3.999996	1.000001	0.000001
5.000000	4.999995	1.000001	0.000001
6.000000	6.000006	0.999999	-0.000001
7.000000	7.000007	0.999999	-0.000001
8.000000	8.000008	0.999999	-0.000001
9.000000	9.000009	0.999999	-0.000001
10.000000	10.00001	0.999999	-0.000001

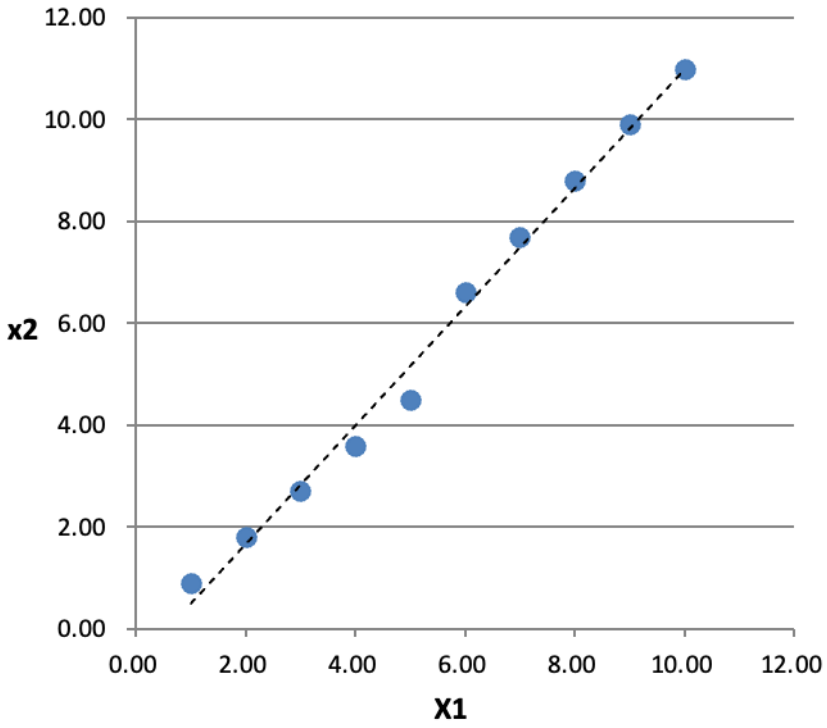


Figure 4-13 Example of associated data (2)

Return now to the Pearson correlation coefficient, the measure most often used to express the degree of association between two measures. In fact, the Pearson correlation coefficient satisfiesK. Tone & Tsutsui (2010), though not rule 4. However, a simple data transformation is all that is needed to adjust its range to [0, 1].

S(a,b)=0.5+0.5R(x1,x2),

R(x1,x2) is the Pearson correlation coefficient of the two measures.

If the projected values of the two measures are radially distributed and their Pearson correlation coefficient is large, this indicates poor substitutability between the two inputs; if the projected values of the two measures show the kind of dispersed distribution seen in Figure 4-11, the Pearson correlation coefficient is negative with a relatively large absolute value, and substitutability between the two inputs is strong.

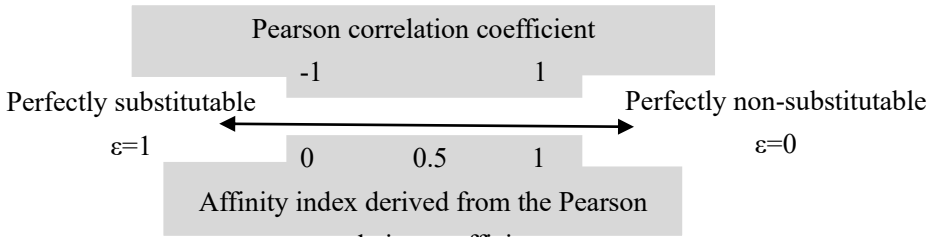


Figure 4-14 Relationship between the Pearson correlation coefficient and the new affinity index

New affinity index computed from Table 4-31

S(P1,P2)=0.5+0.5 R(P1, P2)=0.5+0.5(-1)=0.

Table 4-36 New affinity index matrix

DMU	x1	x2
x1	1	0
x2	0	1

The largest eigenvalue of the new affinity index matrix is 1, so

$$\varepsilon = \frac{m-\max(\rho)}{m-1} = \frac{2-1}{2-1} = 1,$$

the EBM model is equivalent to the SBM model in this case.

Although the analysis above shows that the affinity index computed from the Pearson correlation coefficient is more appropriate for constructing the parameters of the EBM model, in practice the decision maker can determine the values of ε and w entirely from the characteristics of the production technology itself and his or her decision preferences, rather than letting the data determine them.

4.7.3 Solving Hybrid Distance Function Models in MaxDEA

In MaxDEA, apart from the Distance setting, the options for solving a hybrid distance function model are the same as for the other distance function types.

For the EBM model, the method of generating the parameters (ε and w) must be selected:

1) Adopt theK. Tone & Tsutsui (2010) method;

2) Adopt the adjusted Pearson correlation coefficient introduced in this book;

3) parameters entered by the user, including ε and w (measure weights); the ε values for inputs and outputs can be entered directly (Coefficient for Input Nonradial Measure, Coefficient for Output Nonradial Measure), and clicking the "Settings" button opens the interface for setting the weights of the input and output measures.

4.8 Distances That Incorporate Price Information

4.8.1 Cost Efficiency and Revenue Efficiency Models

One reason for the wide application of DEA is that the analysis of technical efficiency does not require price information on inputs and outputs. When price information is available, however, it can be used to analyse allocative efficiency further. If the prices of the inputs are known, for example, cost efficiency can be analysed and allocative efficiency obtained by comparing cost

efficiency with technical efficiency(Farrell, 1957; R. Färe, Grosskopf, & Lovell, 1985). The cost efficiency (CE) model of DMU k under evaluation is expressed as

$$\text{Cost Efficiency} = \frac{cx^*}{cx_k}$$

$$cx^* = \min \ cx$$

$$s.t. \ x \geq X\lambda$$

$$y_k \leq Y\lambda$$

$$\lambda \geq 0 \quad (4.75)$$

c denotes the price vector. The relationship among cost efficiency, allocative efficiency and technical efficiency is

Cost efficiency (CE) = technical efficiency (TE) × allocative efficiency (AE).

When price information on the outputs is known, revenue efficiency can be analysed:

$$\text{Revenue Efficiency} = \frac{ry_k}{ry^*}$$

$$ry^* = \max \ ry$$

$$s.t. \ x_k \geq X\lambda$$

$$y \leq Y\lambda$$

$$\lambda \geq 0 \quad (4.76)$$

When price information on both inputs and outputs is known, profit efficiency can also be analysed:

Table 4-37 Programming formulations of the cost and revenue efficiency models

Cost efficiency	$$\text{Cost Efficiency} = \frac{cx^*}{cx_k}$$ $$cx^* = \min \ cx$$ $$s.t. \ x \geq X\lambda$$ $$y_k \leq Y\lambda$$ $$\lambda \geq 0$$
Revenue efficiency	$$\text{Revenue Efficiency} = \frac{ry_k}{ry^*}$$ $$ry^* = \max \ ry$$ $$s.t. \ x_k \geq X\lambda$$ $$y \leq Y\lambda$$ $$\lambda \geq 0$$

c denotes the input price vector and r denotes the output price vector

4.8.2 Solving Cost Efficiency and Related Models in MaxDEA

To solve cost efficiency and related models, the corresponding price measures must be added to the data table in addition to the input and output measures.

After the data have been imported into MaxDEA, the price measures are left as "Not defined" when the data are defined. Solving a cost efficiency model requires the input prices to be set; solving a revenue efficiency model requires the output prices to be set.

For the model distance, select cost or revenue, then click the settings button on the right and choose the price variables of the input and/or output measures from the drop-down list. Models such as the cost efficiency model involve no choice of model orientation, but the returns-to-scale type can be selected.

MaxDEA does not report allocative efficiency in its results; users can compute it as required, for example by running the input-oriented radial DEA model and the cost model separately and

then taking the ratio of the resulting cost efficiency to technical efficiency, which is the allocative efficiency.

4.9 Analysing the Degree of Inefficiency of Individual Input and Output Measures

Building on DEA efficiency scores and projection analysis, the degree of inefficiency of each individual input and output measure can be analysed further.

The degree of inefficiency of an individual input measure of DMU k is computed as

$\frac{x_{ik}-\hat{x}_{ik}}{x_{ik}}$, $\hat{x}_{ik}$ denotes x_{ik}'s projected value;

The degree of inefficiency of an individual output measure is computed as

$\frac{\hat{y}_{rk}-y_{rk}}{y_{rk}}$, $\hat{y}_{rk}$ denotes y_{rk}'s projected value.

It should be noted that the degree of inefficiency of an individual input or output measure given above pertains only to the distance type used for the current efficiency measurement and to its projection point. Different distance types yield different projections and hence different degrees of inefficiency for individual input and output measures. Even within the same distance type, such as the directional distance function, different direction vectors yield different projections. Consider the directional distance function in the figure below: with different direction vectors, the projection point of G lies on the broken line ABCD. When the projection is point A, the degree of inefficiency of input X2 is 0 and all input inefficiency is concentrated in X1; when the projection is point D, the degree of inefficiency of input X1 is 0 and all input inefficiency is concentrated in X2. The results of an analysis of the degree of inefficiency of individual input and output measures therefore cannot be divorced from the efficiency measurement method adopted and its projection point. It is meaningless to state simply that the degree of inefficiency of a certain input or output measure of a DMU equals a certain value. The analysis of the degree of inefficiency of individual input and output measures should be regarded as a further interpretation of the chosen efficiency measurement method and its projection, not as an evaluation of the extent to which a

certain input (output) is "wasted" ("insufficient") in the production process of the DMU under evaluation.

MaxDEA does not provide results for the analysis of the degree of inefficiency of individual input and output measures; users can compute them from the projected values reported by the software.

Chapter 5 Super-efficiency Models

In the results of a DEA model, several DMUs are often evaluated as efficient; in particular, when there are many input and output measures, the number of efficient DMUs is also large. The efficiency score obtained from a DEA model is at most 1, so all efficient DMUs have the same score and cannot be further distinguished from one another. To solve this problem,Andersen & Petersen (1993) proposed a method that further discriminates the degree of efficiency among efficient DMUs, and this method has since become known as the "super-efficiency" model. For ease of exposition, the models introduced above are called standard efficiency models. The core of the super-efficiency model is to exclude the DMU under evaluation from the reference set; that is, the efficiency of the DMU under evaluation is derived with reference to the frontier formed by the other DMUs, and the super-efficiency score of an efficient DMU is generally greater than 1, so that efficient DMUs can be discriminated.

In many applications of efficiency analysis, the determinants of efficiency must be analysed further. Because an efficiency score of at most 1 is regarded as a censored value, the Tobit regression model is widely used in the literature. Super-efficiency, by contrast, is not subject to the censoring of efficiency scores, so there is no need to use a Tobit regression model designed specifically for censored values.

AlthoughAndersen & Petersen (1993) proposed a radial super-efficiency model, its principle applies equally to other types of distance function. For example,K. Tone (2002) proposed the SBM super-efficiency model,Ray (2008) proposed the directional distance function super-efficiency model, and so on.

As for the ranking function of the super-efficiency model,Banker & Chang (2006) found by data simulation that its ranking of efficient DMUs is not reliable, but this has not prevented the super-efficiency model from being applied ever more widely.

Other common uses of the super-efficiency model include:

1) detecting outliers(Wilson, 1995; Banker & Chang, 2006);

2) sensitivity and stability analysis of DEA models(W. W. Cooper et al., 2001; J. Zhu, 2001; Jahanshahloo, Hosseinzadeh, Shoja, Sanei, & Tohidi, 2005; Boljuncic, 2006; F. Hosseinzadeh Lotfi, Jahanshahloo, & Esmaeili, 2007);

3) generalized DEA (introduced separately below);

4) analysing changes in productivity through intertemporal comparison (introduced separately below).

5.1 Radial Super-efficiency Model

Taking the input-oriented CRS radial model as an example, the only difference between the super-efficiency model and the standard efficiency model is the additional restriction j≠k, that is, the exclusion of the DMU under evaluation, k, from the reference set(Andersen & Petersen, 1993):

$$\begin{aligned}
&\min \theta \\
&\text{s.t.} \sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j x_{ij} \leq \theta x_{ik} \\
&\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j y_{rj} \geq y_{rk} \\
&\lambda \geq 0 \\
&i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k)
\end{aligned} \quad (5.1)$$

The VRS super-efficiency model simply adds the constraint $\sum_{j=1, j\neq k}^{n} \lambda_j = 1$ on the basis of the CRS model:

$$\begin{aligned}
&\min \theta \\
&\text{s.t.} \sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j x_{ij} \leq \theta x_{ik} \\
&\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j y_{rj} \geq y_{rk} \\
&\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j = 1 \\
&\lambda \geq 0 \\
&i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k)
\end{aligned} \quad (5.2)$$

The basic principle of the super-efficiency model is illustrated graphically below. Again taking the data in Table 2-3 used in the graphical illustration of the input-oriented radial standard efficiency model as an example, in the standard efficiency model the four efficient DMUs A, B, C and D form the efficient frontier. Consider the super-efficiency of C: the super-efficiency of C is measured against the frontier formed by the DMUs other than C, namely ABD, and the projection

of C onto that frontier is the point C'.[i] Clearly, C is more efficient than C'; compared with its projection point C', the part by which the efficiency of C exceeds that of C' is reflected in CC', and its super-efficiency score is $\theta^* = OC'/OC \geqslant 1$.

For an efficient DMU, the radial projection direction in the input-oriented super-efficiency model is an increase in inputs. Holding the inputs and outputs of all other DMUs constant, the efficient DMU is still evaluated as efficient after its inputs have been increased proportionally (by a proportion of $\theta^* - 1$); this is the basic principle underlying the use of super-efficiency models in sensitivity and stability analysis.

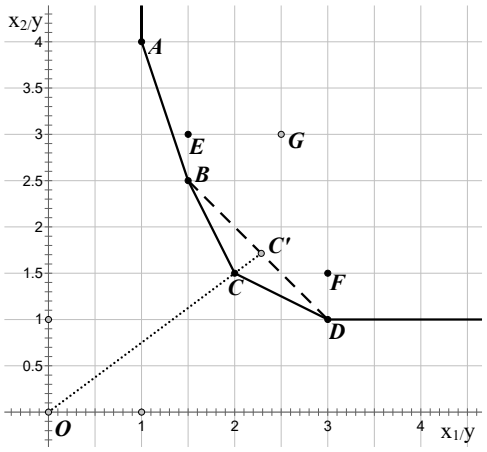


Figure 5-1 Illustration of the Super-efficiency Model

The output-oriented VRS super-efficiency model is expressed as

$$\max \phi$$

$$\text{s.t.} \sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j x_{ij} \leq x_{ik}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j y_{rj} \geq \phi y_{rk}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j = 1$$

$$\lambda \geq 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n\ (j \neq k) \quad (5.3)$$

Although the super-efficiency model is designed for efficient DMUs, it applies equally to

[i]Note that the projected point of a super-efficiency model may also lie on the frontier of the standard efficiency model; for example, if there is another DMU on the frontier between points A and B, the super-efficiency projected point of that DMU is itself, which still lies on the frontier of the standard efficiency model.

inefficient DMUs, and for an inefficient DMU the radial super-efficiency model yields the same result as the radial standard efficiency model.

5.2 Directional Distance Function Super-efficiency Model

Following the same principle as the radial super-efficiency model, the directional distance function super-efficiency model also adds the restriction j≠k to the standard efficiency model(Ray, 2008). Taking the VRS model as an example, the directional distance function super-efficiency model that includes undesirable outputs is expressed as:

$$\max \beta$$

$$\text{s.t.} \sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j x_{ij} + \beta g_{xi} \leq x_{ik}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j y_{rj} - \beta g_{yr} \geq y_{rk}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j b_{tj} - \beta g_{yt} \leq b_{tk}$$

$$\sum_{j=1, j\neq k}^{n} \lambda_j = 1$$

$$\lambda \geq 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q_1; t = 1,2\ldots,q_2; j = 1,2\ldots,n\ (j \neq k) \quad (5.4)$$

As with the radial super-efficiency model, the directional distance function super-efficiency model also applies to inefficient DMUs, and for an inefficient DMU the super-efficiency model yields the same result as the standard efficiency model.

5.3 SBM(ERM) Super-efficiency Model

Compared with the radial and directional distance function super-efficiency models, the SBM(ERM) super-efficiency model is somewhat more complex, since it does not merely add the restriction j≠k(K. Tone, 2002). The programming formulation of SBM(ERM) super-efficiency applies only to efficient DMUs; for an SBM(ERM)-efficient DMU k, taking VRS as an example, the non-oriented SBM(ERM) super-efficiency model is expressed as

$$\min \rho_{\mathrm{SE}} = \frac{\frac{1}{m}\sum_{i=1}^{m} \bar{x}_i \ / x_{ik}}{\frac{1}{q}\sum_{r=1}^{q} \bar{y}_r \ / y_{rk}}$$

$$s.t.\ \ \bar{x}_i \ \geq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j$$

$$\bar{y}_r \ \leq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j$$

$$\bar{x}_i \ \geq x_{ik}$$

$$\bar{y}_r \ \leq y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+, \bar{y} \geq 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n\ (j \neq k) \quad (5.5)$$

The production possibility set constructed from the DMUs other than DMU k is

$$\{(x,y): x \geq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, y \leq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j\}\ .$$

The projected value of the DMU under evaluation, k, in the SBM(ERM) super-efficiency model $(\bar{x}, \bar{y})$; that is, the optimal solution of the model is the point in the production possibility set constructed from the other DMUs that is closest to the frontier. Why is it the point closest to the frontier, rather than a point that necessarily lies on the frontier?

To facilitate comparison with the standard SBM(ERM) efficiency model, the model(5.5) is transformed equivalently as follows:

$$\min \rho_{\mathrm{SE}} = \frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+ \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n\ (j \neq k) \quad (5.6)$$

Why is an equality sign used in the constraints of the standard SBM(ERM) efficiency model, whereas an inequality sign is used in the super-efficiency model?

The SBM(ERM) super-efficiency model yields the minimum distance from the DMU under evaluation to the frontier formed by the other DMUs, so the projection point of the SBM(ERM) super-efficiency model may lie on the weakly efficient part of the frontier. In that case, using an equality sign in the constraints of the SBM(ERM) super-efficiency model may render the model infeasible.

Taking the VRS SBM(ERM) super-efficiency model as an example, if an equality sign is used in the constraints, the model becomes

$$\min \rho_{\mathrm{SE}}=\frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- = x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ = y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+ \geq 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.7)$$

Let $\hat{x}=\sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^-$, $\hat{y}=\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+$; then the feasible region of $(\hat{x},\hat{y})$ is

$$\{(\hat{x},\hat{y}): \hat{x}\leq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, \hat{y}\geq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j, \sum_{j=1,\, j\neq k}^{n} \lambda_j = 1\}.$$

If some input of DMU k is greater than that of any other DMU, or some output is smaller than that of any other DMU, that is,

$x_{ik} > x_{ij}, j\neq k$ or $y_{rk} < y_{rj}, j\neq k$,

then $(x_k, y_k) \notin \{(\hat{x},\hat{y}): \hat{x}\leq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, \hat{y}\geq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j, \sum_{j=1,\, j\neq k}^{n} \lambda_j = 1\}$, that is, the feasible region $(\hat{x},\hat{y})$ does not contain DMU k.

This is because, given $\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$ as the constraint, the expression $x_{ik} > x_{ij}, j\neq k$ or $y_{rk} < y_{rj}, j\neq k$ necessarily implies

$$x_{ik} > \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j \text{ or } y_{ik} < \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j.$$

This shows that if $x_{ik} > x_{ij}, j\neq k$ or $y_{rk} < y_{rj}, j\neq k$, the model(5.7) has no feasible solution.

Taking the single-input, single-output VRS model as an example,Table 5-1 are distributed as shown in Figure 5-2. The value of the input x of C is higher than that of every other DMU; if the model is used(5.7) to compute the super-efficiency of C, $(\hat{x}, \hat{y})$ has the feasible region shown as the shaded area in the figure; C does not lie in this region, so the model(5.7) has no solution. The value of the output y of A is lower than that of every other DMU, and if the model is used(5.7) to compute the super-efficiency of A, no feasible solution is obtained either.

Table 5-1 Sample Data for the SBM(ERM) Super-efficiency Model

DMU	x	y
A	1.0	1.0
B	1.5	2.0
C	4.0	3.0
D	3.5	2.5
E	2.5	1.5

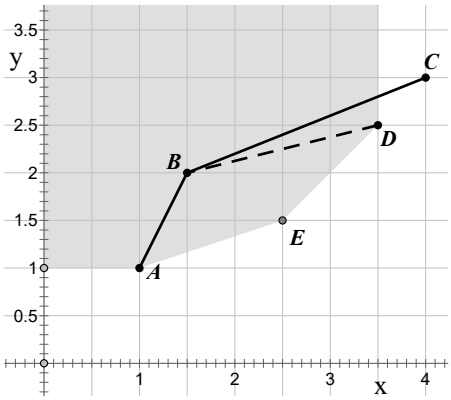


Figure 5-2 Illustration of a Model(5.7) That Has No Feasible Solution

Although the above discussion is based on the VRS model, the problem of infeasibility is not limited to the VRS model: it also arises in the CRS model, although the reason is not as easy to explain as in the VRS case. For example, in the CRS SBM(ERM) super-efficiency model, if an equality sign is used in the constraints,Table 2-3[i].

[i]Note that the next section deals specifically with the infeasibility problem of the VRS super-efficiency model, but this problem is not the same as the one discussed here. The infeasibility problem of the VRS super-efficiency model introduced later exists not only in the input-oriented and output-oriented SBM super-efficiency models but also in the input-oriented and output-oriented radial super-efficiency models and directional distance function super-efficiency models.

The SBM(ERM) super-efficiency model(5.6) above is the non-oriented SBM(ERM) super-efficiency model, and the input-oriented and output-oriented SBM(ERM) super-efficiency models are expressed respectively as

$$\min \rho_{\mathrm{SE}} = 1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}$$

$$s.t. \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j \ge y_{rk}$$

$$\lambda, s^-, s^+ \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.8)$$

$$\min \rho_{\mathrm{SE}} = \frac{1}{1-\frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

$$s.t. \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\lambda, s^-, s^+ \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.9)$$

The super-efficiency SBM(ERM) model yields the minimum distance from the DMU under evaluation to the frontier formed by the other DMUs, and its projection point is not necessarily strongly efficient; the super-efficiency SBM(ERM) model can therefore also be regarded as the super-efficiency solution model for the MinDS and MinDW models.

The SBM(ERM) super-efficiency model with undesirable outputs (VRS) is

$$\min \rho_{\mathrm{SE}} = \frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk} + \sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right)}$$

$$s.t. \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} b_{tj}\lambda_j - s_t^{b-} \le b_{tk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+ \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q_1; t=1,2\ldots,q_2; j=1,2\ldots,n\ (j\neq k) \quad (5.10)$$

5.4 Weighted Additive Distance Super-efficiency Model

The weighted additive distance super-efficiency model is similar to the SBM(ERM) super-efficiency model and is therefore not described in detail here.

Taking the VRS model as an example, if DMU k is an efficient DMU (that is, its inefficiency score = 0), its weighted additive distance super-efficiency model is

$$\max \ -\sum_{i=1}^{m} w_i^- s_i^- - \sum_{r=1}^{q} w_r^+ s_r^+$$

$$s.t. \ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j b_{tj} - s_t^{b-} \le b_{tk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+ \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q_1; t = 1,2\ldots,q_2; j = 1,2\ldots,n \ (j \neq k) \quad (5.11)$$

The inefficiency score obtained from the weighted additive distance super-efficiency model is negative.

5.5 Projection Analysis in Super-efficiency Models

Chapter 2 of this book (Section 2.12, Common Misunderstandings about Projected Values) explained the meaning of the projected values of a DEA model. The projected value (target value) is the projection of the DMU under evaluation onto the frontier, and the relative gap between the current value of a DMU and its projected value represents the degree of inefficiency of each of its inputs and outputs.

In the standard efficiency model, the projected values of an inefficient DMU are always better than its original values, that is, the projected input is $\leqslant$ the original input, the projected output is $\geqslant$ the original output, and the projected undesirable output is $\leqslant$ the original undesirable output.

In the super-efficiency model, the results for inefficient DMUs (including their projected values) are exactly the same as the results in the standard efficiency model. For an efficient DMU (efficiency score $\geqslant$ 1), however, there is no fixed ordering between its projected values and its original values. In the super-efficiency model, the efficient DMU under evaluation is projected onto the frontier formed by the other DMUs; since the DMU under evaluation lies in front of that frontier, the relationship between its projected values and its original values is rather complex and depends on the distance function and on the model orientation. A simple summary table is provided here, and interested readers may analyse the matter further.

Table 5-2 Relationship between the Projected and Original Values of Efficient DMUs (Efficiency Score ≥ 1) in Super-efficiency Models

Distance type	Input-oriented	Output-oriented	Non-oriented
Radial distance and directional distance function[i]	Relationship between projected and original input is indeterminate (usually projected input $\geqslant$ original input) Projected desirable output $\geqslant$	Projected input $\leqslant$ original value Relationship between projected and original outputs is indeterminate	Relationship between projected and original input is indeterminate (usually projected input $\geqslant$ original input) Relationship between projected and

[i]Taking the non-oriented super-efficiency model as an example, the gap between the projected value and the target value consists of two parts, a proportional change component and a slack change component. The two have opposite signs (directions of change). Which of the two, the proportional change component or the slack component, has the larger absolute value determines the magnitude relationship between the projected value and the target value.

	original value Projected undesirable output ⩽ original value	(usually projected desirable output ⩽ original value, projected undesirable output ⩾ original value)	original outputs is indeterminate (usually projected desirable output ⩽ original value, projected undesirable output ⩾ original value)
Farthest distance to the frontier (ERM, SBM)	Projected input ⩾ original value Projected desirable output ⩾ original value Projected undesirable output ⩽ original value	Projected input ⩽ original value Projected desirable output ⩽ original value Projected undesirable output ⩾ original value	Projected input ⩾ original value Projected desirable output ⩽ original value Projected undesirable output ⩾ original value
Weighted additive distance			Projected input ⩾ original value Projected desirable output ⩽ original value Projected undesirable output ⩾ original value

5.6 Detecting Outliers with Super-efficiency Models

The non-parametric frontier nature of DEA implies that if an outlier appears on the frontier, a large number of anomalous results may arise, and even the results for all other DMUs may become anomalous. Confirming whether outliers are present is therefore an important part of the analysis. When the data set is large, it is very difficult to detect outliers by direct observation.

An important function of the super-efficiency model is the detection of outliers. In what follows, the data in Table 2-6 are modified to anomalous values, and the outliers are then identified with the super-efficiency model. The input data of Fujian are reduced to 10% of their original values, with all other data left unchanged. A comparison of the results before and after the change is given in Table 5-3. After the values for Anhui are changed, almost all DMUs change their efficiency scores in the results of the input-oriented model. The change is particularly pronounced in the CRS model: apart from Anhui itself, the efficiency scores of all other DMUs

fall to between 1/10 and 1/6 of their original values. The change in the VRS model is relatively small: the results for 4 efficient DMUs are unaffected, while the efficiency scores of the other DMUs decline continuously, the smallest change being only 4% and the largest falling to close to 1/10 of the original value. In terms of super-efficiency, both the CRS and the VRS super-efficiency scores of Fujian are greater than 10, which is consistent with the proportion by which its data were changed.

This simple example shows that the super-efficiency model is an effective way of detecting outliers. In practice, possible outliers can be detected by varying the distance function, the model orientation and the RTS type and by examining several super-efficiency models.

Table 5-3 Detecting Outliers with Super-efficiency Models (Input-oriented)

	CRS			VRS		
Region	**Normal Efficiency score**	**Outlier Efficiency score**	**Outlier Super-efficiency score**	**Normal Efficiency score**	**Outlier Efficiency score**	**Outlier Super-efficiency score**
Anhui	0.9442	0.1006	0.1006	0.9549	0.5786	0.5786
Beijing	0.9174	0.1487	0.1487	0.9185	0.4967	0.4967
Fujian	1.0000	**1.0000**	**10.4553**	**1.0000**	**1.0000**	**10.5133**
Gansu	0.8324	0.0979	0.0979	0.8538	0.1458	0.1458
Guang dong	1.0000	0.1490	0.1490	1.0000	1.0000	1.3051

Chapter 6 Other Models and Methods

6.1 The Multiplier Ratio Constraint Model

In the multiplier model, if additional constraints are imposed on the ratios of input and output weights, the resulting model is called a restricted multiplier model, also known as a cone ratio model, a constrained cone model or an assurance region model.

6.1.1 Constraining the Weight Ratio of Two Inputs

The constraint that sets upper and lower bounds on the ratio of the weight of one input to the weight of another input is expressed as

$$R_L \le \frac{v_a}{v_b} \le R_U \quad (6.1)$$

Economically, this ratio is the marginal rate of technical substitution (MRTS) of input a for input b, that is, the ratio of the increase in the quantity of one factor of production a to the quantity of another factor b that can be reduced while output is held constant. On an isoquant, the ratio represents the slope of the tangent line at a point. Since the factors of production used in a production process are imperfectly substitutable, as the quantity of one factor increases, its marginal rate of technical substitution for the other factor diminishes.

When the researcher knows the quantitative marginal substitution relationship between two inputs, adding such constraints makes the analysis results more reasonable. For example, adding weight ratio constraints can solve the problem of zero weights appearing in the analysis results (for the zero-weight problem, see Section 2.13.3 of Chapter 2). When the researcher lacks knowledge of this quantitative relationship, the upper and lower bounds of the ratio constraint can be set to the maximum and minimum of the corresponding weight ratios in the analysis results for the sample under study. For example, in Figure 6-1, it can be observed that the frontier consists of the four vertices A, B, C and D, and that the projection H' of H lies on the line through A parallel to the horizontal axis (the weak efficient frontier), so that the weight obtained for X1 is necessarily 0. The maximum of the ratio of the weight of input X1 to that of input X2 (v1/v2) occurs on segment CD, where the ratio (that is, the absolute value of the slope) is 2; the minimum of the weight ratio occurs on segment AB, where the ratio is 0.5. If the ratio of the weight of X1 to that of X2 is restricted to the interval 0.5 to 2.0, that is,

$$0.5 \le \frac{v_1}{v_2} \le 2,$$

the projection of H becomes H'', which effectively resolves the problems of a zero weight for X1 and of a weakly efficient projection point.

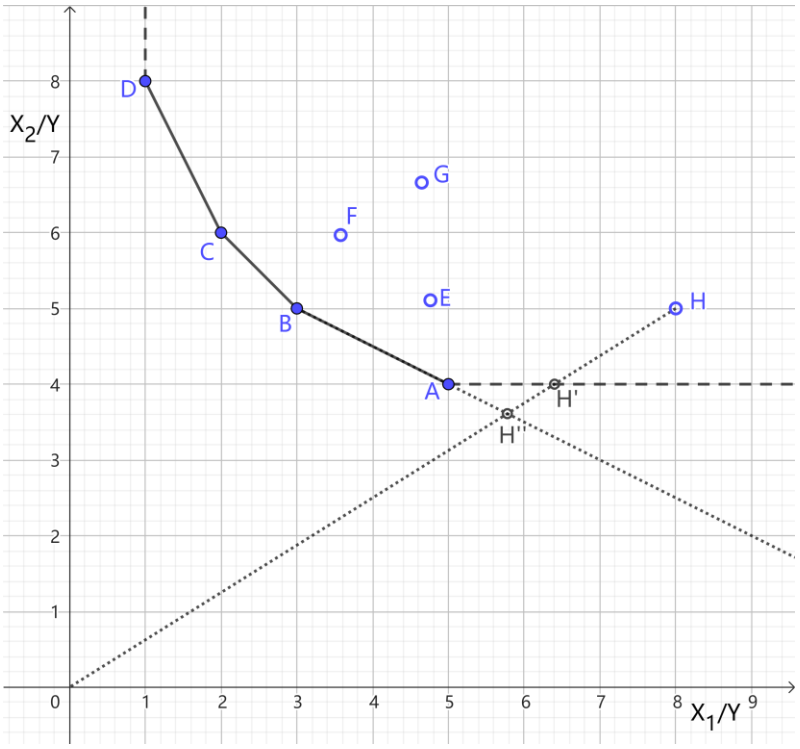


Figure 6-1 Example of Weight Ratio Constraints (2-Input, 1-Output CRS Input-Oriented Model)

6.1.2 Constraining the Weight Ratio of an Output and an Input

$$R_L \le \frac{v_a}{u_b} \le R_U \quad (6.2)$$

When there is only one input and one output, the economic meaning of this ratio is the marginal productivity of x, that is, the ratio of the increase in output to the increase in the input factor; on the production frontier curve it represents the slope of the tangent line, or the derivative with respect to input x. When there are multiple inputs and multiple outputs, the ratio is the partial derivative of that output with respect to that input, and its meaning is the marginal productivity of that input for that output, with the other inputs and outputs held constant. In Figure 6-2, it can be observed that the frontier consists of the four vertices A, B, C and D, and that the projection H' of H lies on the line through A parallel to the vertical axis (the weak efficient frontier), so that the

weight obtained for Y is necessarily 0. The maximum of the ratio of the weight of input X to the weight of output Y (v/u) occurs on segment AB, where the ratio (that is, the slope) is 2; the minimum of the weight ratio occurs on segment CD, where the ratio is 0.5. If the ratio of the weight of X to the weight of Y is restricted to the interval 0.5 to 2.0, that is,

$$0.5 \le \frac{v}{u} \le 2,$$

the projection of H becomes H", which effectively resolves the problems of a zero weight for Y and of a weakly efficient projection point.

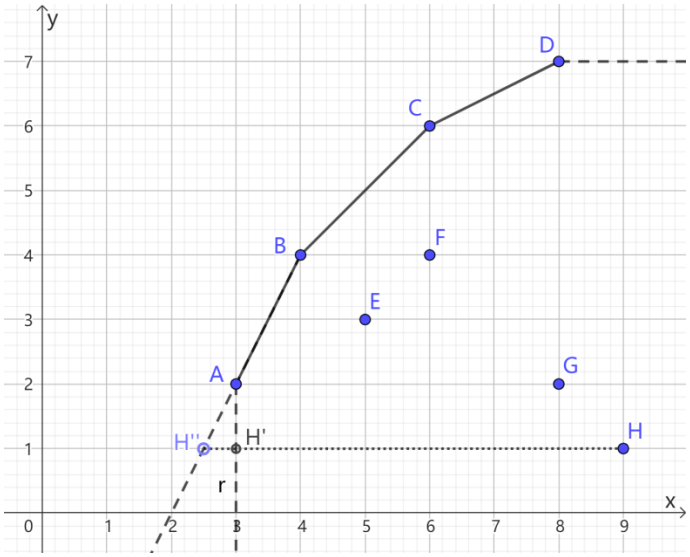


Figure 6-2 Example of Weight Ratio Constraints (1-Input, 1-Output VRS Input-Oriented Model)

6.1.3 Constraining the Weight Ratio of One Output and Another Output

The constraint that sets upper and lower bounds on the ratio of the weight of one output to the weight of another output is expressed as

$$R_L \le \frac{u_a}{u_b} \le R_U \quad (6.3)$$

Its economic meaning is analogous to that of the marginal rate of technical substitution: the ratio of the reduction in the quantity of one product to the increase that can be achieved in the quantity of another product while input quantities are held constant. Imposing output weight ratio constraints likewise resolves the problem of zero weights in the results, and this is not repeated here.

6.2 The FDH Model

The FDH model was proposed byTulkens (1993) as a mixed integer linear programming (MILP) model. Judged from its programming formulation, the FDH model can be regarded as the VRS DEA model with the linear combination coefficients λ restricted to be either 0 or 1, that is, $\lambda \in \{0,1\}$. The full name of the FDH model is Free Disposal Hull, whose literal meaning is a freely disposable hull (envelope). The name FDH mainly reflects the meaning of free disposal, that is, the free disposability (strong disposability) of inputs and outputs in the production possibility set.

Taking the input-oriented model as an example, the FDH model is expressed as

$$\min \theta$$

$$s.t. \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{ik}$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{rk}$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda \in \{0,1\}$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \qquad (6.4)$$

Because of the constraint $\lambda \in \{0,1\}$, the frontier of the FDH model is no longer a convex hull, and its shape is shown in Figure 6-3, where the dashed line is the VRS DEA frontier and the solid line is the FDH frontier.

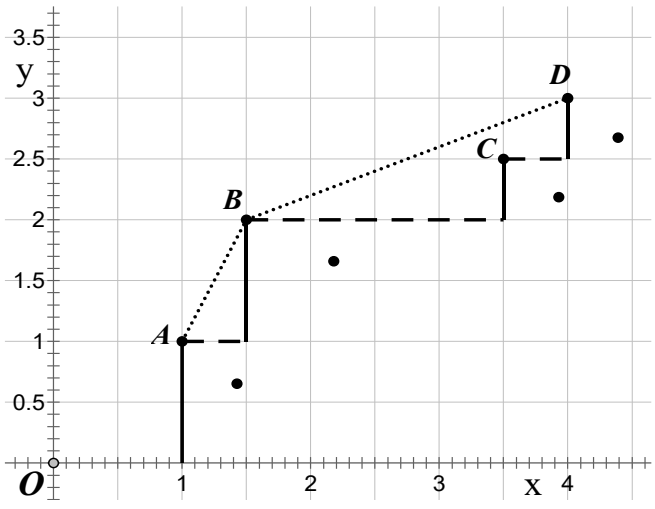


Figure 6-3 Schematic Diagram of the FDH Model (Input-Oriented)

It can be seen from the figure that the production possibility set of the FDH model is a subset of that of the VRS DEA model. In the FDH model, the benchmark of the evaluated DMU can only be a single DMU, with a linear combination coefficient $\lambda = 1$. The significance is that the evaluated DMU refers only to DMUs that actually exist in reality and never to a virtual DMU formed by a linear combination; this is the practical value of the FDH model.

The FDH model can be extended from the input-oriented radial model above to output-oriented and non-oriented models, as well as to other distance types.

6.3 Models Containing Undesirable Outputs

How undesirable output measures are handled in DEA models has been described in detail in the preceding part on the directional distance function. For an understanding of the results of DEA models containing undesirable outputs, see "4.5.8 Misunderstandings That Easily Arise about the Results of DEA Models Containing Undesirable Outputs" in this book.

The treatment of undesirable outputs in the various distance functions is summarised once more here.

6.3.1 Radial Model Containing Undesirable Outputs

$$\begin{aligned}
&\max\ \beta \\
&s.t.\ X\lambda+\beta x_k \le x_k \\
&\quad Y\lambda-\beta y_k \ge y_k \\
&\quad B\lambda-\beta b_k \le b_k \\
&\quad \lambda \ge 0 \quad (6.5)
\end{aligned}$$

6.3.2 SBM Model Containing Undesirable Outputs

The SBM model containing undesirable outputs as defined by Tone is(William W Cooper et al., 2007):

$$\min\ \rho=\frac{1-\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1+\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk}+\sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right)}$$

$$\begin{aligned}
&s.t.\ X\lambda+s^- = x_k \\
&\quad Y\lambda-s^+ = y_k \\
&\quad \mathrm{B}\lambda+s^{b-} = b_k \\
&\quad \lambda, s^-, s^+ \ge 0 \quad (6.6)
\end{aligned}$$

In the preceding discussion, this book recommends setting undesirable outputs as strongly disposable in the directional distance function or the radial model. The programming formulation of the SBM shows that its production possibility set is indeed based on strong disposability of

undesirable outputs.

Tone did not give the programming formulation of the SBM super-efficiency model containing undesirable outputs, but from the SBM super-efficiency model**(5.6)** the SBM super-efficiency model containing undesirable outputs can be derived as

$$\min \ \rho = \frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk} + \sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right)}$$

$$s.t. \ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} b_{tj}\lambda_j - s_t^{b-} \le b_{tk}$$

$$1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk} + \sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right) > 0$$

$$\lambda, s^-, s^+ \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \ (j \neq k) \quad (6.7)$$

where the additional constraint $1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk} + \sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right) > 0$ can be removed in the linear transformation.

In DEA models, undesirable outputs can also be handled by expressing them as negative numbers. In addition, MaxDEA provides no dedicated model for undesirable inputs; this can be dealt with by converting undesirable input measures into negative numbers, as described later in the introduction to negative measures.

MaxDEA supports negative measures and allows various model options to be combined. For the SBM super-efficiency model containing undesirable outputs, for instance, it suffices to select the SBM model, tick SuperEfficiency and Undesirable Outputs on the Advanced Models(1) interface and make the corresponding settings.

6.3.3 Directional Distance Function Model Containing Undesirable Outputs

$$\max \ \beta$$

$$s.t. \ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda-\beta g_b \le b_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0, g_b \le 0 \quad (6.8)$$

6.3.4 Weighted Additive Model Containing Undesirable Outputs

6.3.4.1 Simple Additive Model Containing Undesirable Outputs

$$\max \ \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{q_1} s_r^+ + \sum_{t=1}^{q_2} s_t^{b-}$$

$$s.t. \ X\lambda+s^- = x_k$$

$$Y\lambda-s^+ = y_k$$

$$\mathrm{B}\lambda+s^{b-} = b_k$$

$$\lambda, s^-, s^+ \ge 0 \quad (6.9)$$

6.3.4.2 Weighted Additive Model Containing Undesirable Outputs

$$\max \ \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q_1} w_r^+ s_r^+ + \sum_{t=1}^{q_2} w_t^{b-} s_t^{b-}$$

$$s.t. \ X\lambda+s^- = x_k$$

$$Y\lambda-s^+ = y_k$$

$$\mathrm{B}\lambda+s^{b-} = b_k$$

$$\lambda, s^-, s^+ \ge 0 \quad (6.10)$$

6.3.5 Handling Undesirable Outputs by Means of Negative Measures

Some DEA models allow negative values among input and output measures, for example the directional distance function model with a user-defined direction vector, the RDM directional distance function model and the MSBM model.

Radial, SBM and MinDS models: input-oriented models allow negative values in output data; output-oriented models allow negative values in input data; non-oriented models do not allow negative values.[i]

The radial model is taken as an example to explain why negative values are not allowed. The input-oriented radial VRS model can be equivalently expressed as

$$\begin{aligned} &\min 1-\beta \\ &s.t.\ X\lambda+\beta x_k \le x_k \\ &\quad Y\lambda \ge y_k \\ &\quad e\lambda=1 \\ &\quad \lambda \ge 0 \end{aligned} \quad (6.11)$$

$x_{ik}-\alpha x_{ik}$ is the target value of the input measure of the evaluated DMU after radial improvement, and the improvement is $-\alpha x_{ik}$, a negative number indicating that the direction of improvement is a reduction in input. When the input measure xk is negative, the improvement becomes positive and the direction of radial improvement becomes a proportional increase in input.

Similarly, in an output-oriented model, when the output measure yk is negative, the erroneous result also arises that the direction of improvement becomes a proportional reduction in output.

In the radial model, if in the improvement term $-\alpha x_k$ the absolute value of x_k is taken, the above logical error can be avoided; moreover, if the input measures are positive, the modified model is equivalent to the original model(Cheng, Zervopoulos, & Qian, 2013). The modified radial model (variant of radial measure, VRM) is expressed as

[i]Some scholars hold that the above allowance for negative numbers applies only to the VRS model, because only the VRS model satisfies translation invariance: when a positive number is added to all measures that contain negative values, the results of the model remain unchanged, so that negative measures can be converted into positive measures by translation. In the author's view, translation invariance is a sufficient rather than a necessary condition for allowing negative numbers. Whether the above allowance for negative numbers can be used with other RTS types depends on whether the technology and the data under study exhibit the characteristics of the corresponding RTS type. If the technology is such that inputs become λ times their original level (Xλ) and outputs, including negative outputs, also become λ times their original level (which is characteristic of CRS), then the above allowance for negative numbers can be applied to the CRS model (for example, using negative numbers to represent undesirable outputs).

$$\min 1-\beta$$

$$s.t.\ X\lambda+\beta|x_k|\le x_k$$

$$Y\lambda\ge y_k$$

$$e\lambda=1$$

$$\lambda\ge 0 \qquad (6.12)$$

In the above formulas, | | denotes taking the absolute value of the variable. When no negative numbers are involved, that is, when xk is positive, the optimal solution of the modified DEA linear programme is equivalent to that of the traditional DEA linear programme; in other words, the new DEA model remains compatible with the traditional DEA model. When yk is negative, taking the absolute value ensures the correct direction of improvement for inefficiency.

The modified output-oriented VRM model is expressed as

$$\min\ \frac{1}{1+\beta}$$

$$s.t.\ X\lambda\le x_k$$

$$Y\lambda-\beta|y_k|\ge y_k$$

$$e\lambda=1$$

$$\lambda\ge 0 \qquad (6.13)$$

The SBM model can be treated in a similar way, except that only its objective function needs to be modified:

$$\min\ \rho=\frac{1-\frac{1}{m}\sum_{i=1}^{m}s_i^-/|x_{ik}|}{1+\frac{1}{q}\sum_{r=1}^{q}s_r^+/|y_{rk}|}$$

$$s.t.\ X\lambda+s^-=x_k$$

$$Y\lambda-s^+=y_k$$

$$e\lambda=1$$

$$\lambda,s^-,s^+\ge 0 \qquad (6.14)$$

The VRM model has two problems:

1) When the input or output measures contain both negative and positive values, the efficiency scores obtained from the modified model do not satisfy monotonicity; that is, it is possible that after the evaluated DMU reduces its inputs or increases its outputs, the resulting

efficiency score falls instead. In this case, a measure containing both positive and negative values can be decomposed into two measures, holding the positive part and the negative part of the measure separately:[i]

$$v_{ij}^1 = \begin{cases} v_{ij} & if\ v_{ij} \geq 0, \\ 0 & if\ v_{ij} < 0, \end{cases} \text{ and } v_{ij}^2 = \begin{cases} 0 & if\ v_{ij} \geq 0, \\ v_{ij} & if\ v_{ij} < 0. \end{cases} \quad (6.15)$$

2) The VRM model can be applied to CRS only under certain conditions. The CRS VRM model applies only when the negative measures contain negative values exclusively and, as the scale of production expands (inputs increase), the absolute values of the negative measures increase. For example, when undesirable outputs are expressed as negative numbers, undesirable outputs increase accordingly as the scale of production expands.

If undesirable outputs are expressed as negative numbers, the VRM model is equivalent to the radial model containing undesirable outputs described above. The VRM model containing negative output measures (output-oriented) is

$$\begin{aligned} &\min \ \frac{1}{1+\beta} \\ &s.t.\ X\lambda \leq x_k \\ &\quad Y\lambda - \beta |y_k| \geq y_k \\ &\quad \lambda \geq 0 \quad (6.16) \end{aligned}$$

where the constraint $Y\lambda - \beta |y_k| \geq y_k$ can be decomposed into constraints on the positive output measures and on the negative output measures:

[i]This method of separating positive and negative numbers draws onEmrouznejad, Anouze, & Thanassoulis (2010) the SORM (semi-oriented radial measure) method, but with differences.Emrouznejad et al. (2010) method, both variables after the separation are kept positive, that is, $v_{ij}^1 = \begin{cases} v_{ij} & if\ v_{ij} \geq 0, \\ 0 & if\ v_{ij} < 0, \end{cases}$ and $v_{ij}^2 = \begin{cases} 0 & if\ v_{ij} \geq 0, \\ -v_{ij} & if\ v_{ij} < 0. \end{cases}$.

$$Y^{+}\lambda-\beta y_k \geq y_k^{+},$$

$$Y^{-}\lambda-\beta(-y_k) \geq y_k^{-}.$$

Multiplying both sides of the constraints on the negative output measures by -1,

$$(-Y^{-})\lambda+\beta(-y_k) \leq (-y_k^{-}),$$

the resulting model is equivalent to the output-oriented radial model containing undesirable outputs**(4.42)**.

Similarly, undesirable inputs such as garbage and wastewater treatment can be handled by means of negative measures.

6.4 Handling Negative Values in MaxDEA

Some DEA models support negative values in themselves, for example the directional distance function, the weighted additive distance and the MSBM model. When the selected model does not itself support negative values, MaxDEA automatically applies the VRM described above (or a similar treatment) so that the selected model admits negative values.

6.5 External Uncontrollable Factors or Uncontrollable Input/Output Measures

In practical applications of efficiency or productivity analysis, it is often the case that the production process is affected by external factors that decision makers cannot control; the efficiency of public health services, for example, is affected by the convenience of transport and the population density in the service area. If such factors can be reduced to numerical variables and converted into positive-direction measures, that is, measures for which a larger value indicates a more favourable production environment, they can be treated in the DEA model as uncontrollable (non-discretionary) input measures. In production practice there are also cases in which the quantity of some inputs (outputs) is not determined or controlled by the decision maker; these can likewise be treated as uncontrollable input (output) measures. For example, the quantity of medical services provided by a hospital is not something the hospital's decision makers can determine directly, although they can attract patients by improving the quality of service.

Taking the DEA model containing uncontrollable inputs as an example, uncontrollable inputs and discretionary inputs are treated differently in the model(Banker & Morey, 1986). The VRS model is taken as an example; its linear programme is:

$$\min \theta$$

$$s.t.\ X_D\lambda - \theta x_{\mathrm{k},D} \geq 0$$

$$X_N\lambda \leq x_{\mathrm{k},N}$$

$$Y\lambda \geq y_k$$

$$e\lambda = 1$$

$$\lambda \geq 0 \qquad (6.17)$$

D denotes discretionary inputs and N denotes non-discretionary inputs.

In the constraints, the uncontrollable inputs do not enter into a direct relationship with the efficiency score θ, so that their role is confined to ensuring the reasonableness of the benchmarking (the production environment of the projection point is not more favourable than that of the evaluated DMU); the principle is similar to the control of confounding factors in statistical analysis.

In the literature, different scholars have named this class of DEA models differently. The model was first proposed byBanker & Morey (1986), who called it a DEA model with exogenously fixed inputs and outputs beyond the discretionary control of DMU managers;J. Zhu (2009) termed it a measure specific model;William W Cooper et al. (2007) termed it a nondiscretionary variable model.

In MaxDEA, to build a model containing uncontrollable inputs/outputs it is only necessary, when selecting the DEA model and distance function, to tick the adjacent "Uncontrollable Inputs/Outputs" box and then click "Settings", whereupon the input and output measures that are uncontrollable variables can be specified.

Chapter 7 Generalized DEA Models

Starting from this chapter, the content no longer concerns the options of the DEA models themselves, such as the types of distance function. Both the generalized DEA of this chapter and the panel data DEA of the next chapter address the relationship between the evaluated set and the reference set in DEA models.

7.1 The Generalized DEA Model

马占新 (2002) proposed the concept of sample DEA, and 马占新 & 马生昀 (2009) proposed, on the basis of sample DEA, the concept of generalized DEA (Generalized DEA).[i]

The core of the generalized DEA model is the separation of the "evaluated unit set" from the "reference set". Traditional DEA evaluates the relative efficiency among decision making units, and the reference set is the evaluated unit set itself; generalized DEA, by contrast, compares the evaluated units with a reference set, and the relationship between the reference set and the evaluated unit set may be inclusion, equality, intersection or disjointness.[ii]

In terms of the concept of the production possibility set, the production possibility set of traditional DEA is formed by the evaluated set, whereas the production possibility set of generalized DEA is formed by the reference set, so that the evaluated objects are separated from the reference standard. One example is the comparison of the decision making unit sets of different groups with the same fixed reference set.

Let V denote the evaluated set (the eValuated set) and T the reference set (the referred Technology or reference set). In generalized DEA, the relationship between the evaluated set ($k \in V$) and the reference set ($j \in T$) includes the following cases:

1) $V=T$ means that the evaluated set is identical to the reference set, that is, traditional

[i] In the sample DEA model proposed by 马占新 (2002), the reference set is called the sample set. 马占新 (2002) proposed "sample DEA" and 马占新 & 马生昀 (2009) proposed "generalized DEA", both of which have at their core the separation of the "set of decision making units" from the "reference set", or in other words the separation of the objects evaluated from the standards of comparison. In terms of naming, "generalized DEA" is more appropriate.

[ii] Cook, Seiford, & Zhu (2004) variable-benchmark DEA model also has at its core the separation of the "set of decision making units" from the "reference set"; in this senseCook et al. (2004) variable-benchmark DEA and the 马占新 (2002) and 马占新 & 马生昀 (2009) sample DEA and generalized DEA are based on the same concept, but 马占新 & 马生昀 (2009) have developed the content of generalized DEA more fully.

DEA;

2) $V \subset T$ means that the evaluated set is a subset of the reference set;

3) $\mathrm{T} \subset \mathrm{V}$ means that the reference set is a subset of the evaluated set;

4) $V \bigcap T = \varnothing$ means that neither the reference set nor the evaluated set contains the other, and the two do not intersect;

5) $V \bigcap T \neq \varnothing$ means that the reference set and the evaluated set intersect.

Taking the input-oriented CRS radial model as an example, the generalized DEA model can be expressed simply as

$$\min \theta$$

$$s.t. \sum_{j \in T} \lambda_j x_{ij} \leq \theta x_{ik}$$

$$\sum_{j \in T} \lambda_j y_{rj} \geq y_{rk}$$

$$\lambda \geq 0$$

$$i = 1, 2 \ldots, m; r = 1, 2 \ldots, q; j \in T, i \in V \quad (7.1)$$

T denotes the reference set, which represents the technology used for benchmarking.

Although generalized DEA was proposed on the basis of radial DEA models, this idea can be extended to other types of distance function, such as the directional distance function, the SBM (ERM), MinDS, MinDW, cost, revenue, profit and other models.

Considered from the perspective of an individual evaluated DMU, if the evaluated DMU is not in the reference set, that is, $k \notin T$, and is also not within the production possibility set of reference set T, that is, $\mathrm{k} = (\mathrm{x_k}, \mathrm{y_k}) \notin \{(\mathrm{x,y}): x \geq \sum_{j \in T} \lambda_j x_{ij}, y \leq \sum_{j \in T} \lambda_j y_{rj}\}$, then generalized DEA is in essence a super-efficiency model.

It should be noted that for some distance functions the programming formulation of the standard efficiency model is not the same as that of the super-efficiency model, in which case the generalized DEA expressions have to be given separately. Taking the SBM (ERM) model as an example, its generalized DEA model can be expressed as

$$\min \rho = \frac{1-\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1+\frac{1}{s}\sum_{r=1}^{s} s_r^+ / y_{rk}}$$

$$s.t. \sum_{j\in \mathrm{T}}^{n} x_{ij}\lambda_j + s_i^- = x_{ik}$$

$$\sum_{j\in \mathrm{T}}^{n} y_{rj}\lambda_j - s_r^+ = y_{rk}$$

$$\lambda, s^-, s^+ \geq 0$$

$$i = 1, 2 \ldots, m;\ r = 1, 2 \ldots, q;\ j \in \mathrm{T},\ i \in V \quad (7.2)$$

If DMU K has no feasible solution in the model, it is then(7.2) necessary to solve the model

$$\min \rho_{\mathrm{SE}} = \frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{s}\sum_{r=1}^{s} s_r^+ / y_{rk}}$$

$$s.t. \sum_{j\in \mathrm{T}}^{n} x_{ij}\lambda_j - s_i^- \leq x_{ik}$$

$$\sum_{j\in \mathrm{T}}^{n} y_{rj}\lambda_j + s_r^+ \geq y_{rk}$$

$$\lambda, s^-, s^+ \geq 0$$

$$i = 1, 2 \ldots, m;\ r = 1, 2 \ldots, q;\ j \in \mathrm{T},\ i \in V \quad (7.3)$$

7.2 Cluster Benchmarking Models

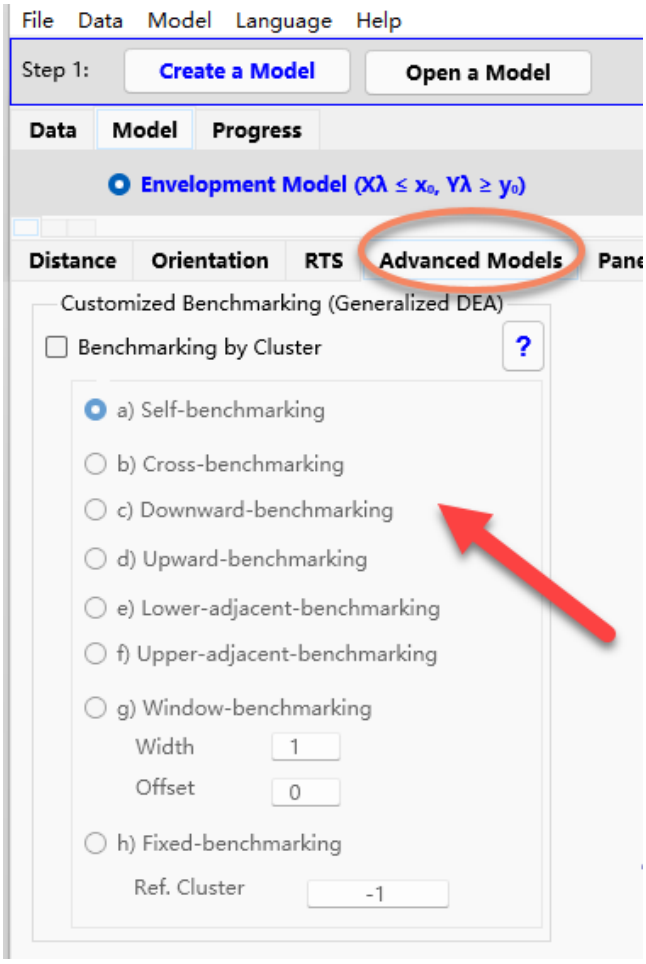


Figure 7-1 Cluster Benchmarking Model

The DMUs in the data can be divided into several groups according to some attribute, for example, the 31 provinces (municipalities and autonomous regions) of mainland China divided into three groups according to the eastern, central and western regions; or the existing data may be processed in some way to generate DMUs of different groups, for example, several new DMU groups generated by means of the frontier-shifting method in generalized DEA. Then, between different groups and according to the different roles of the "evaluated set" and the "reference set", various forms of benchmarking (Cluster Benchmarking) can be carried out.

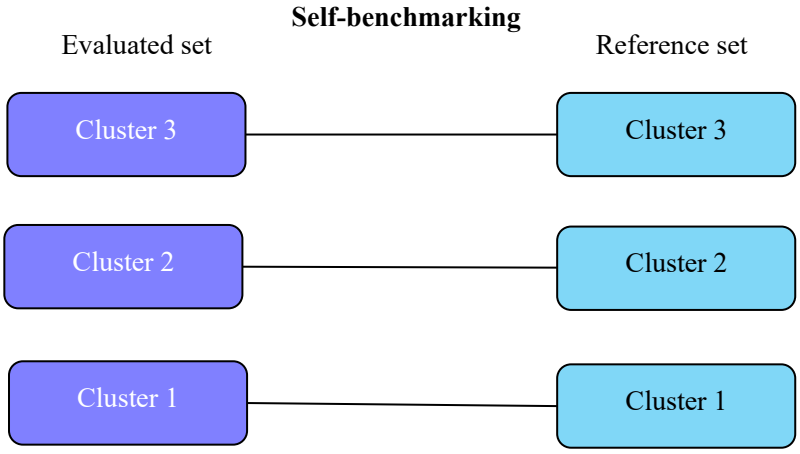
In the following figures, one line represents a pair of an evaluated set and a reference set, and the brace on the right denotes the union of the corresponding sets.

7.2.1 Self-benchmarking

Each group takes itself as the "reference set" and carries out "self-evaluation" separately. If the data of each group are saved in a different data table and each data table is then analysed

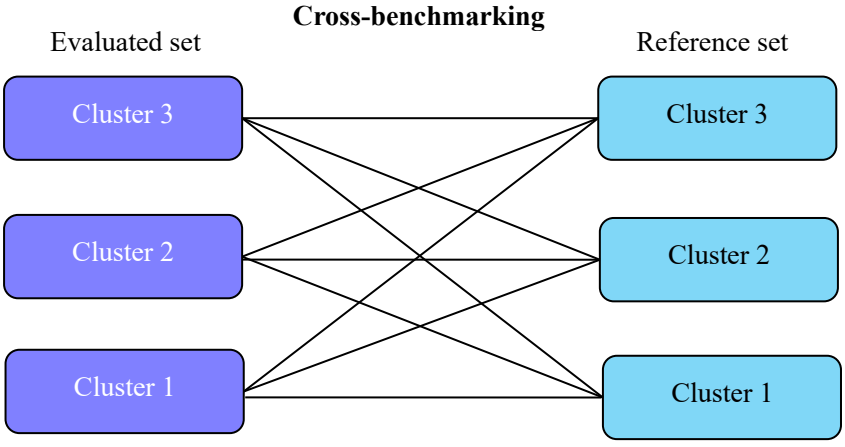
separately, the results obtained are the same as those of self-benchmarking.

The self-benchmarking method makes it convenient to conduct MetaFrontier-DEA (metafrontier DEA) analysis. By comparing the results of the DEA model without grouping (that is, the traditional approach, in which all DMUs form one reference set) with the self-benchmarking results after the groups are defined, the technology gap ratio (TGR) can be obtained.



7.2.2 Cross-benchmarking

Each group serves in turn as the reference set, and the groups are evaluated one after another. Cross-benchmarking includes self-benchmarking.

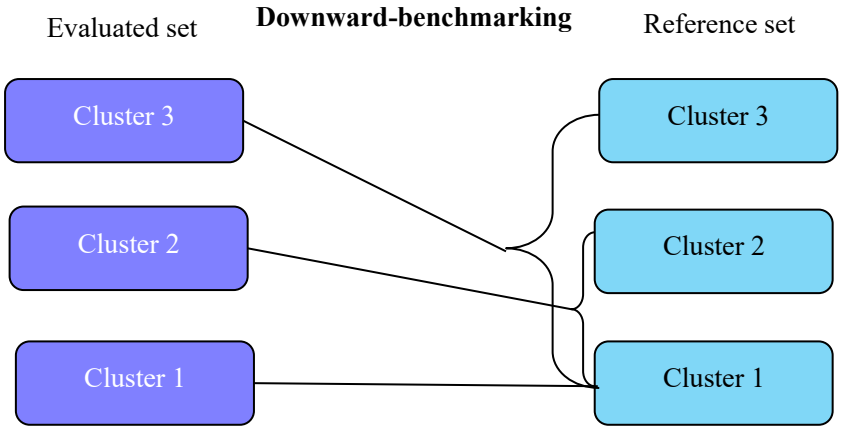


7.2.3 Downward-benchmarking

In downward-benchmarking, the reference set of each group consists of the group itself plus all groups whose group number is smaller than its own. Taking Cluster 3 as an example, its reference set is the union of three sets: Cluster 3 ∪ Cluster 2 ∪ Cluster 1. The reference set of

the group with the smallest group number is the group itself.

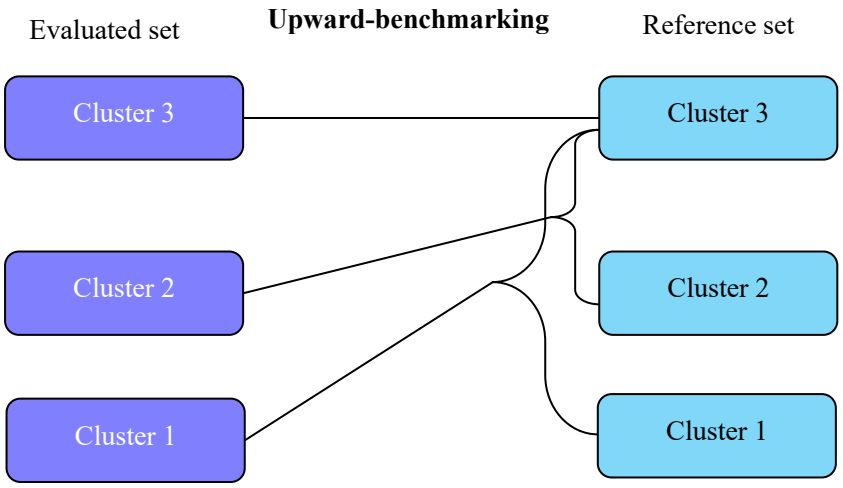
The downward-benchmarking method is applicable when all DMUs are divided into several groups according to the magnitude of the influence of external factors on production efficiency (the greater the negative influence, the lower the group number). The reference set of each group is the group itself plus the groups subject to a greater external influence, that is, the production environment of the reference set is no better than that of the evaluated set.



7.2.4 Upward-benchmarking

In contrast to downward-benchmarking, in upward-benchmarking the reference set of each group consists of the group itself plus all groups whose group number is larger than its own. Taking Cluster 1 as an example, its reference set is the union of three sets: Cluster 3 ∪ Cluster 2 ∪ Cluster 1. The reference set of the group with the largest group number is the group itself.

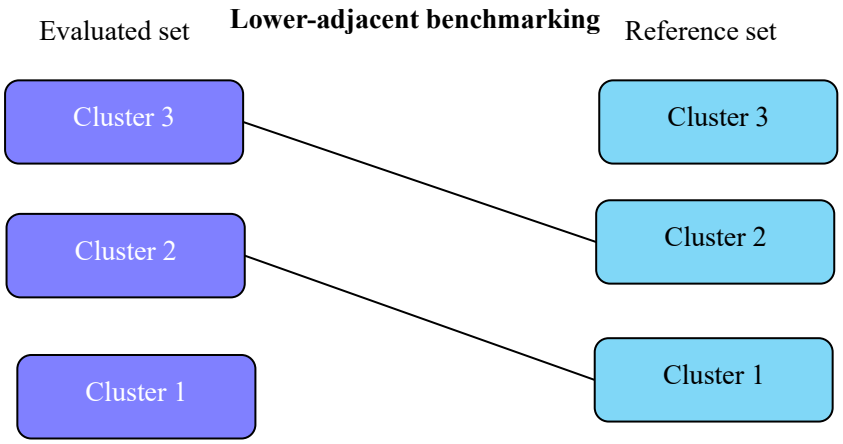
Upward-benchmarking likewise applies when all DMUs are divided into several groups according to the magnitude of the influence of external factors on production efficiency (note that the greater the negative influence, the higher the group number). The reference set of each group is the group itself plus the groups subject to a greater external influence, that is, the production environment of the reference set is no better than that of the evaluated set.



7.2.5 Lower-adjacent Benchmarking

Lower-adjacent benchmarking means that each group takes the one adjacent group whose number is smaller than its own as the reference set. For example, if there are three groups in total, the reference set of Cluster 3 is Cluster 2; the reference set of Cluster 2 is Cluster 1; and Cluster 1 does not serve as an evaluated group.

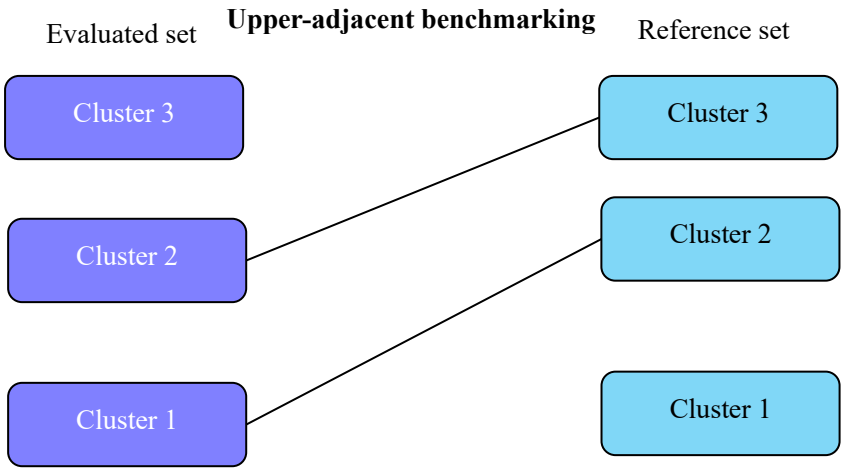
This method can be used to obtain analytical details that are not provided in the analysis results of the panel data Malmquist model; see the introduction to the Malmquist model below.



7.2.6 Upper-adjacent Benchmarking

Upper-adjacent benchmarking means that each group takes the one adjacent group whose number is larger than its own as the reference set. For example, if there are three groups in total, the reference set of Cluster 1 is Cluster 2; the reference set of Cluster 2 is Cluster 3; and Cluster 3 does not serve as an evaluated group.

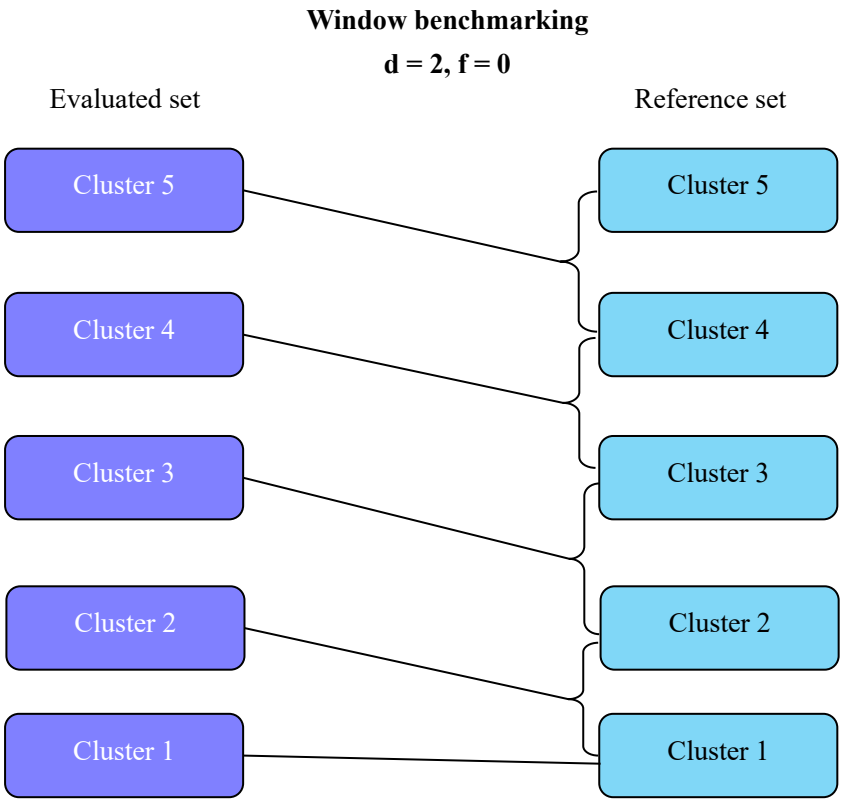
The function of this method is likewise to obtain analytical details that are not provided in the analysis results of the panel data Malmquist model; see the introduction to the Malmquist model below.



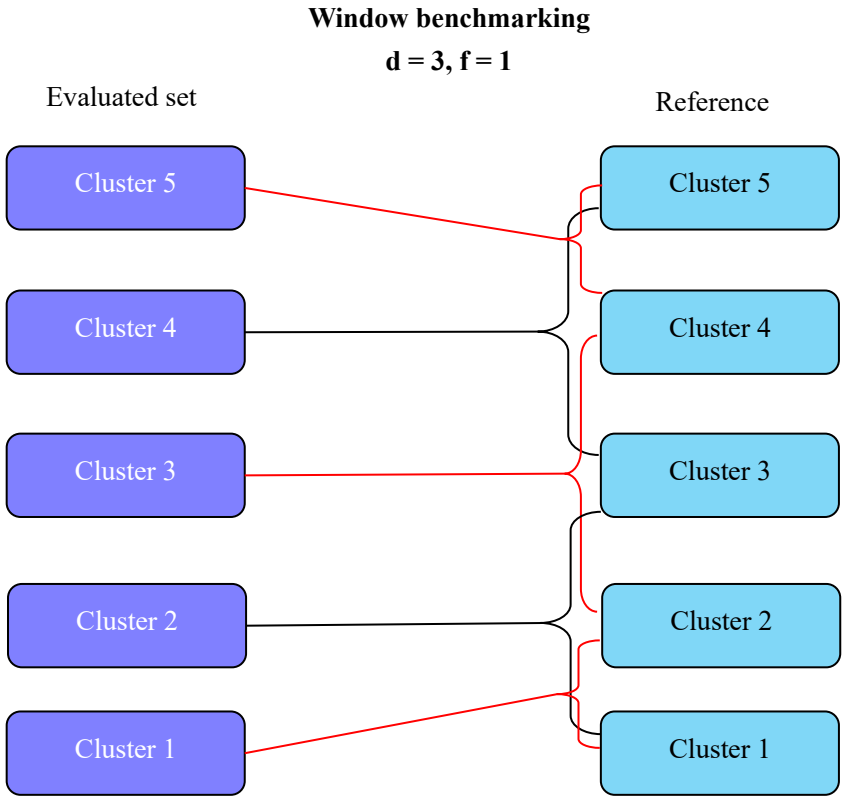
7.2.7 Window Benchmarking

In window benchmarking, the reference set of each group is the union of a series of groups with consecutive numbers; the reference set contains the evaluated group itself, and the maximum number of groups contained in the reference set is determined by the window width (width, denoted by d). By default the window consists of the group in which the evaluated DMU lies and the groups below it. The window can be shifted upward by setting an offset. Suppose that the number of the group in which the evaluated DMU lies is c; after the offset (offset, denoted by f) is set, the range of groups included in the reference set is [c-d+1+f, c+f].

Suppose that there are five groups in total, with a window width of d = 2 and an offset of f = 0; the reference sets of the groups are as shown in the figure below. Note that the reference sets of the evaluated groups whose numbers are close to the two ends (the beginning and the end) are incomplete windows, and the number of incomplete windows is determined by the window width.



If the window width is d = 3 and the offset is f = 1, the reference sets of the groups are as follows:



The function of window benchmarking is likewise to obtain analytical details that are not provided in the analysis results of the panel data Malmquist model; see the introduction to the Window-Malmquist model below.

7.2.8 Fixed Benchmarking

Fixed benchmarking designates a particular group as the reference set.

7.3 Data Format for Cluster Benchmarking Models

For group data, a variable indicating the group number must be included, and its values must be integers, for example, 1, 2, 3,...; they may, however, be non-consecutive, for example, 1, 2, 5, 8,...; 2001, 2005, 2009,...; 200101, 200102, 200302,... and so on.

When group data are imported, the group number variable must be set as Cluster on the data definition interface.

The following is an example of group data:

Table 7-1 Example of Group Cross-sectional Data

DMU	Cluster	Input	Output
A	1	4323	93608
B	1	2295	225559
C	1	6379	327068
D	1	6644	201354
E	1	1436	188926
F	1	6281	413738
G	2	7459	114022
H	2	4464	212444
I	2	4524	462677

Group data may be unbalanced, that is, the number of DMUs contained in each group may differ. For group panel data, each DMU must belong to the same group in different periods. For example, if DMU A belongs to group 1 in period 1, it must belong to group 1 in all other periods.

Table 7-2 Example of Group Panel Data

Period	DMU	Cluster	Input	Output
1	A	1	4323	93608
1	B	1	6644	201354
1	C	1	7459	114022
1	D	1	2295	225559
1	E	1	1436	188926
1	F	1	4464	212444
1	G	2	6379	327068
1	H	2	6281	413738

1	**I**	**2**	4524	462677
2	**A**	**1**	4341	93656
2	**B**	**1**	6689	201442
2	**C**	**1**	7559	114107
2	**D**	**1**	2362	225604
2	**E**	**1**	1513	188937
2	**F**	**1**	4544	212446
2	**G**	**2**	6406	327102
2	**H**	**2**	6369	413797
2	**I**	**2**	4563	462739

Chapter 8 Panel Data Models

8.1 Panel Data

Panel data refer to data drawn from two or more periods. The most common panel data are annual data, that is, each DMU has one set of input and output data in every year. Panel data may be consecutive annual data or data for discontinuous periods, for example, panel data consisting of two periods, the first period being 2000 and the second period being 2020 (Table 8-1). Never add the year to the DMU name; doing so is wrong (Table 8-2).

Table 8-1 Example of Panel Data Format

Period	DMU	Capital	Labor	Product
2000	A	4323	875	93608
2000	B	2295	469	225559
2000	C	6379	1286	327068
2020	A	6644	1339	201354
2020	B	1436	297	188926
2020	C	6281	1266	413738

Table 8-2 Example of an Incorrect Panel Data Format

Period	DMU	Capital	Labor	Product
2000	A2000	4323	875	93608
2000	B2000	2295	469	225559
2000	C2000	6379	1286	327068
2020	A2020	6644	1339	201354
2020	B2020	1436	297	188926
2020	C2020	6281	1266	413738

The DEA models introduced above all use cross-sectional data for a specific period and analyse the technical efficiency of each DMU in that specific period, that is, relative to the production technology of that period. Production, however, is a long-term and continuous process, and in the course of it the production technology itself often changes continuously (technological progress).

When the data of the evaluated DMUs are panel data containing observations at several points in time, it becomes possible to analyse the change in productivity and the respective contributions of technical efficiency and technological progress to that change. This is the widely used Malmquist total factor productivity (TFP) index analysis.

The following simple example shows why panel data cannot be analysed in terms of technical efficiency alone. For the two DMUs A and B, in period 1 the technical efficiency of A (x=4, y=1) is 1 and that of B (x=6, y=1) is 0.667; in period 2 the technical efficiencies of A (x=1, y=4) and B (x=1.5, y=4) do not change, A still being 1 and B still 0.667. If only technical efficiency is analysed, then from period 1 to period 2 the technical efficiencies of A and B do not change at all. In reality, however, the output-input ratios (that is, the productivities) of both A and B increase substantially: compared with period 1, the productivity of both A and B in period 2 rises 15-fold (productivity in period 2 is 16 times that in period 1). This is the effect of technological progress.

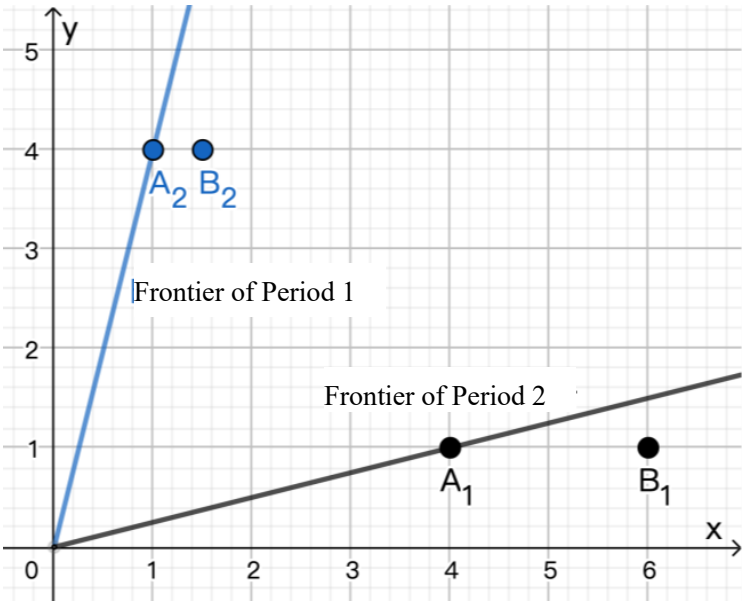


Figure 8-1 Example of Panel Data Analysis

In addition, window DEA is also a commonly used method of panel data analysis: it takes all DMUs within a time span of a certain width (the window period) as a single reference set, thereby multiplying the number of data points in the reference set (a data point may be regarded as a "DMU"). Window DEA is an effective way of dealing with an insufficient number of DMUs(A. Charnes, Clark, Cooper, & Golany, 1985).

MaxDEA allows unbalanced panel data to be used, and the results corresponding to missing data are missing accordingly. For unbalanced panel data, a warning message is displayed once the run is complete, and the log file gives a more detailed explanation.

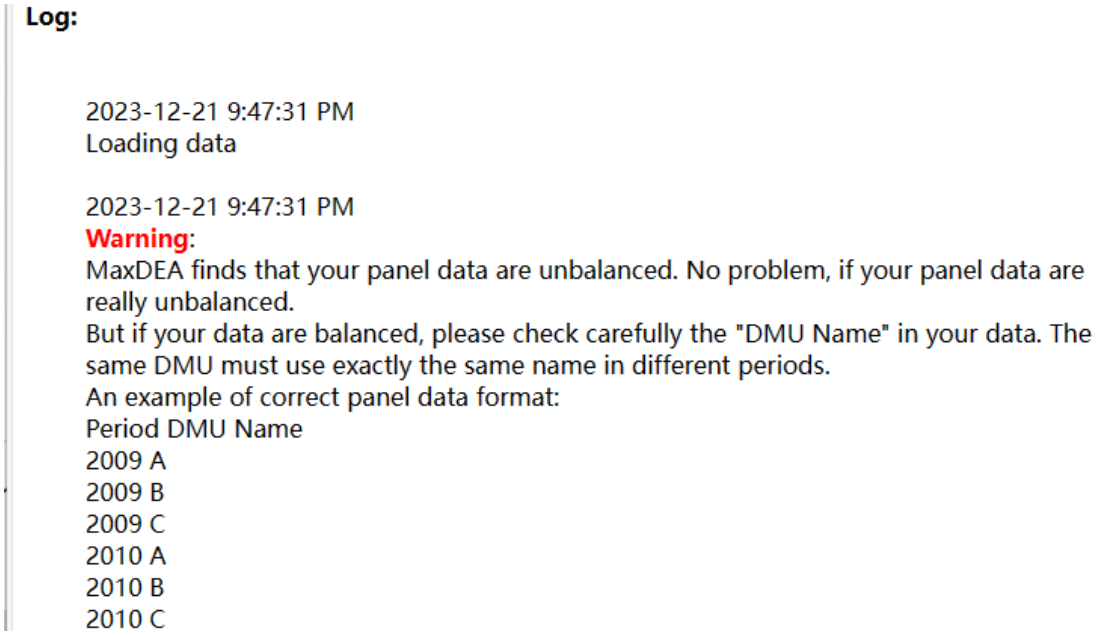


Figure 8-2 MaxDEA Warning on the Use of Unbalanced Panel Data

Balanced panel data mean that each DMU has one and only one observation in every period. If a DMU has two or more observations in the same period, for example, if in Table 8-3Table 8-3 there are two DMUs named "A" in 2020, the software reports an error (Figure 8-3).

Table 8-3 Example of Incorrect Panel Data

Period	DMU	Capital	Labor	Product
2000	A	4323	875	93608
2000	B	2295	469	225559
2000	C	6379	1286	327068
2020	**A**	6644	1339	201354
2020	**A**	1436	297	188926
2020	C	6281	1266	413738

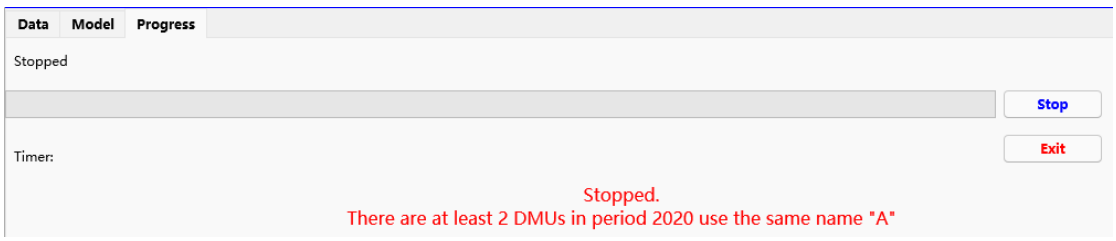


Figure 8-3 MaxDEA Warning on Incorrect Panel Data

计数项:Province	列标签									
行标签	2003	2004	2005	2006	2007	2008	2009	2010	2011	总计
P01	1	1	1	1	1	1	1	1	1	9
P02	1	1	1	1	1	1	1	1	1	9
P03	1	1	1	1	1	1	1	1	1	9
P04	1	1	1	1	1	1	1	1	1	9
P05	1	1	1	1	1	1	1	1	1	9
P06	1	1	1	1	1	1	1	1	1	9
P07	1	1	1	1	1	1	1	1	1	9
P08	1	1	1	1	1	1	1	1	1	9
P09	1	1	1	1	1	1	1	1	1	9
P10	1	1	1	1	1	1	1	1	1	9
P11	1	1	1	1	1	1	1	1	1	9
P12	1	1	1	1	1	1	1	1	1	9
P13	1	1	1	1	1	1	1	1	1	9
P14	1	1	1	1	1	1	1	1	1	9
P15	1	1	1	1	1	1	1	1	1	9
P16	1	1	1	1	1	1	1	1	1	9
P17	1	1	1	1	1	1	1	1	1	9
P18	1	1	1	1	1	1	1	1	1	9
P19	1	1	1	1	1	1	1	1	1	9
P20	1	1	1	1	1	1	1	1	1	9
P21	1	1	1	1	1	1	1	1	1	9
P22	1	1	1	1	1	1	1	1	1	9
P23	1	1	1	1	1	1	1	1	1	9
P24	1	1	1	1	1	1	1	1	1	9
P25	1	1	1	1	1	1	1	1	1	9
P26	1	1	1	1	1	1	1	1	1	9
P27	1	1	1	1	1	1	1	1	1	9
P28	1	1	1	1	1	1	1	1	1	9
P29	1	1	1	1	1	1	1	1	1	9
P30	1	1	1	1	1	1	1	1	1	9
P31	1	1	1	1	1	1	1	1	1	9
总计	31	31	31	31	31	31	31	31	31	279

Figure 8-4 Checking Whether Panel Data Are Balanced in Excel

If the message MaxDEA gives for an unbalanced panel is in doubt, the PivotTable function in Excel can be used to check whether the panel data are balanced: insert a PivotTable in Excel, use the DMU name variable as the row label and the count item, and use the period variable as the column label. If every value in the crosstab is 1 and there are no blank entries, the panel is balanced.

If the panel data are very large, checking whether they are balanced in Excel is not easy either, and statistical software can be used instead.

The Stata command for checking whether panel data are balanced (Stata calls this strongly balanced) is:

```
xtset dmu_name_variable period_variable
spbalance
```

xtset requires both variables to be numeric; if the DMU name variable is a string variable, first create a new id variable:

```
egen id = group(dmu_name_variable)
xtset id period_variable
```

spbalance

When the volume of data is large, whether the data are balanced cannot be judged correctly by eye. When organising data, never rely on the eye.

Is seeing believing? Look at the figure below: your eyes tell you that the colours of the two regions A and B are different, yet the two colours are in fact exactly the same.

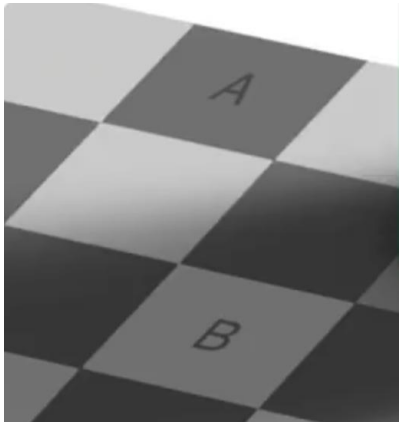


Figure 8-5 Optical Illusion

8.2 The Relationship between Productivity Change, Technological Change and Technical Efficiency Change

When input-output analysis is carried out on two periods (t-1, t) of panel data, three concepts are involved: productivity change, technological change and technical efficiency change. To make the relationship among the three easier to grasp, several special cases are considered first.

(1) Figure 8-6: technological progress in the industry (outward shift of the frontier), with the productivity of the evaluated DMU K unchanged in the two periods (t-1, t) (input and output values unchanged)

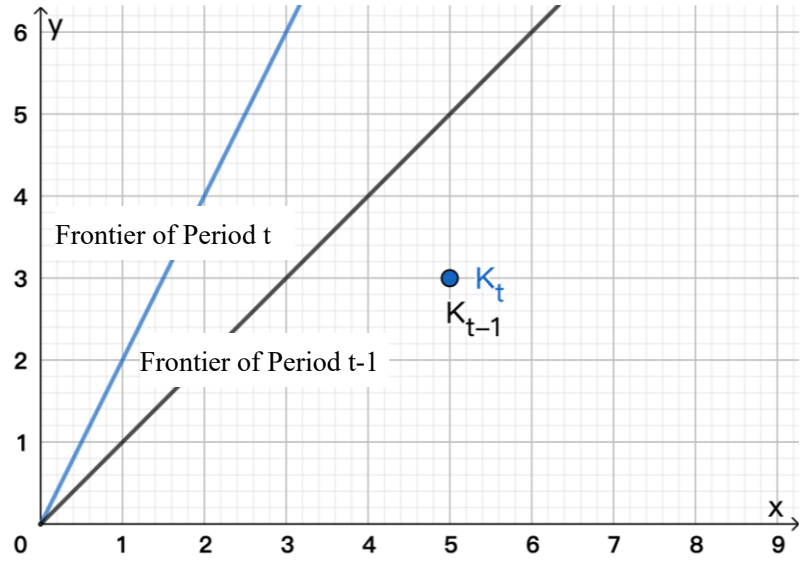


Figure 8-6 Relationship between Productivity Change, Technological Change and Technical Efficiency Change (1)

Productivity is the ratio of output to input. Since neither the input nor the output of K changes, the productivity of K does not change either: its productivity is the same in the two periods, that is, the productivity index is 1. Because the frontier shifts outward in period t, technological progress occurs in period t relative to period t-1, and the technological change index is 2. In period t the distance from K to the frontier becomes larger, so its technical efficiency falls and the technical efficiency index is 0.5 (technical efficiency is 0.3 in period t and 0.6 in period t-1, a ratio of 0.5).

It should be noted that technological change refers to the production technology of the industry in a given period, not to the production technology adopted by a particular DMU.

In this example the production situation of K does not change at all in the two periods. Although the productivity of K remains unchanged, technological progress in the industry (higher frontier productivity) reduces the technical efficiency of K (its relative distance to the frontier becomes larger).

For a production process with multiple inputs and multiple outputs, although the DEA method cannot compute the productivity of K in the two periods, it can compute the productivity index of K in period t relative to period t-1. In practice, the numerical value of productivity is not necessarily required; what matters is the change in productivity, that is, the productivity index.

(2) Figure 8-7: industry technology unchanged, with the productivity of the evaluated DMU K rising in period t (input unchanged but output increased)

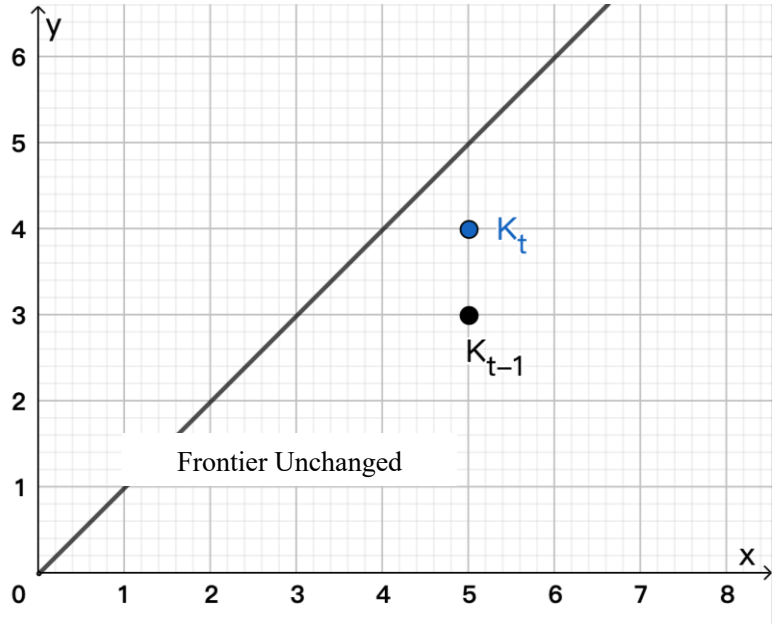


Figure 8-7 Relationship between Productivity Change, Technological Change and Technical Efficiency Change (2)

The productivity of K rises in period t, with a productivity index of 1.333. Industry technology does not change (the frontier remains unchanged). In period t the distance from K to the frontier becomes smaller, so its technical efficiency rises, that is, the technical efficiency index is 1.333 (technical efficiency is 0.8 in period t and 0.6 in period t-1, a ratio of 1.333).

(3) Figure 8-8: technological progress in the industry, with the productivity of the evaluated DMU K rising in period t (input unchanged but output increased)

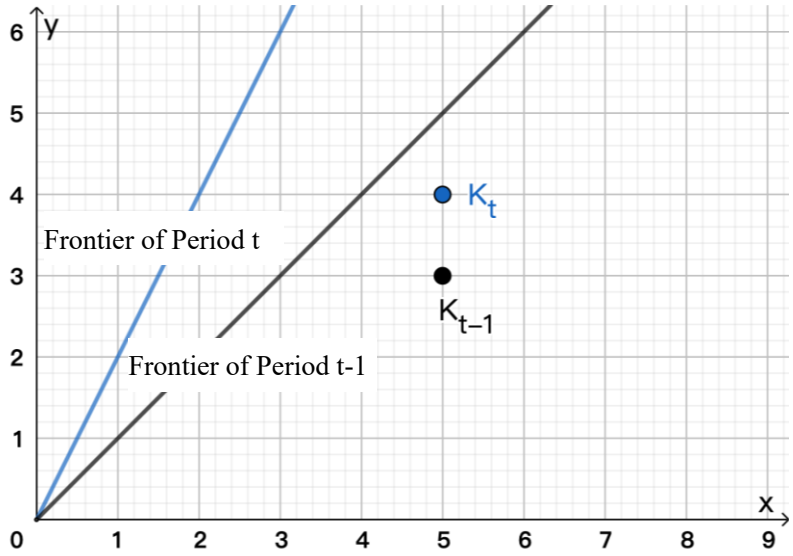


Figure 8-8 Relationship between Productivity Change, Technological Change and Technical Efficiency Change (3)

The productivity of K rises in period t, with a productivity index of 1.333. Because the frontier shifts outward in period t, that is, technological progress occurs in period t relative to period t-1, the technological change index is 2. In period t the distance from K to the frontier becomes larger, so its technical efficiency falls and the technical efficiency index is 0.667 (technical efficiency is 0.4 in period t and 0.6 in period t-1, a ratio of 0.667).

8.3 Basic Principles of the Malmquist Index

Panel data are data involving two or more periods, and every DMU has input and output values for several periods. Technology changes over time, which is what is usually called technological progress, and technological progress is often the primary driving force of productivity growth. If the panel data are split into single-period data and each period is analysed separately, the technical efficiency value for each year can be obtained. This yields only the technical efficiency values of the individual periods; when technological progress is the main driving force of productivity growth, analysing technical efficiency alone is not comprehensive.

Total factor productivity (TFP) refers to the part of the quantity of output that cannot be attributed to the quantities of the factors of production. TFP is easier to understand from a dynamic perspective. Suppose that, with the factors of production of a production unit unchanged, its output increases by 30% compared with the previous year. This 30% increase in output cannot be attributed to the quantities of the factors of production but is caused by factors other than the factors of production, such as technological progress; this is the role of TFP. The TFP of the production unit is 30% higher than in the previous year (the TFP index is 1.3).

The DEA method is a non-parametric method and cannot compute TFP, but it can compute a TFP index. Suppose that the input x of a DMU remains unchanged in periods t and t-1 and that its outputs are yt and yt-1 respectively; then

$$y_{t-1}=TFP_{t-1}\times f(x) \quad (8.1)$$

$$y_{t}=TFP_{t}\times f(x) \quad (8.2)$$

Since the production function f(x) is unknown, the TFP of the two periods cannot be computed, but its TFP index can be computed; this is the basic idea of the DEA method for computing a TFP index.

$$\text{TFP Index (t-1,t)}=\frac{TFP_t}{TFP_{t-1}}=\frac{y_t}{y_{t-1}} \quad (8.3)$$

The concept of the Malmquist total factor productivity index (MI) originated withMalmquist (1953), hence this class of indices is named the Malmquist index.R Färe, Grosskopf, Lindgren, & Roos (1992) were the first to use the DEA method to compute the Malmquist index, and decomposed the Malmquist index into two components of change: the change in technical efficiency (technical efficiency change, EC) and the change in production technology (production frontier) (technological change, TC).

Different types of Malmquist index differ only in the frontier used as reference. Which distance type, which orientation and which more complex DEA model options are adopted in computing the Malmquist index is a matter of the DEA model itself; see the relevant chapters above. The choice of DEA model and the choice of Malmquist index type are two mutually independent questions. To keep the explanation of the computation of the Malmquist index as simple as possible, the simplest DEA model is used for the graphical illustration, namely the single-input single-output CRS model.

For the input and output data of the evaluated DMU K in the two periods, what is of interest is its TFP in the two periods, that is, TFPt and TFPt-1. When production involves multiple inputs and multiple outputs, TFP cannot be computed, but the TFP index, that is, MI(t-1, t), can be computed. The basic idea of computing MI(t-1, t) by the DEA method is as follows:

$$\text{MI}\ (\text{t-1, t}) = \frac{\text{TFP}(\text{x}_\text{t}, \text{y}_\text{t})}{\text{TFP}(\text{x}_\text{t-1}, \text{y}_\text{t-1})} = \frac{\text{TFP}(\text{x}_\text{t}, \text{y}_\text{t}) \ / \ \text{TFP(benchmark)}}{\text{TFP}(\text{x}_\text{t-1}, \text{y}_\text{t-1}) \ / \ \text{TFP(benchmark)}} \quad (8.4)$$

It was explained in Chapter 2 (2.2.2, 2.2.4, 2.3.2, 2.3.4) that the efficiency value obtained from a DEA model is in essence the ratio of the productivity of the evaluated DMU to the frontier productivity. Computing MI(t-1, t) therefore reduces to computing the ratio of two DEA efficiency values, the numerator being the DEA efficiency value of DMU K in period t and the denominator the DEA efficiency value of DMU K in period t-1; the two efficiency values must, however, be computed against the same frontier.

Both the numerator and the denominator use the frontier of period t-1 as the reference (Figure 8-9), $\text{TFP}(\text{x}_\text{t}, \text{y}_\text{t}) \ / \ \text{TFP}(\text{benchmark}_\text{t-1})$ is the DEA efficiency value of (xt, yt) obtained against the frontier of period t-1, denoted Score_t-1(x_t, y_t);[i] $\text{TFP}(\text{x}_\text{t-1}, \text{y}_\text{t-1}) \ / \ \text{TFP}(\text{benchmark}_\text{t-1})$ is the DEA efficiency value of (xt-1, yt-1) obtained against the frontier of period t-1, denoted Score_t-1(x_t-1, y_t-1).

$$\text{MI}^{\text{t-1}}\text{(t-1,t)} = \frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})} \quad (8.5)$$

[i]The output of MaxDEA is in plain text format, so superscripts and subscripts cannot be used. To make the software output easier to understand, the expressions in this book are consistent with the software output. Readers writing their own papers are advised to follow the expressions used in the original references. In addition, both this book and MaxDEA use (t, t-1) to denote two adjacent periods, which differs from the (t+1, t) notation used in other literature only in the symbols employed.

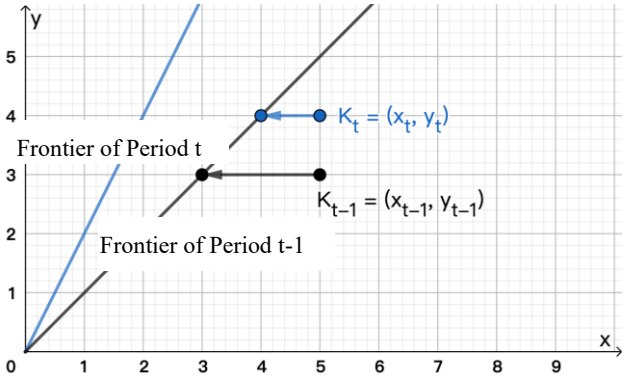


Figure 8-9 Schematic Diagram of the Malmquist Index (with the Frontier of Period t-1 as Reference, CRS)

This is the MI(t-1, t) obtained against the frontier of period t-1. Similarly, against the frontier of period t (Figure 8-10) another MI(t-1, t) can be computed.

$$MI^{t}(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t(x_t\text{-}1, y_t\text{-}1)} \quad (8.6)$$

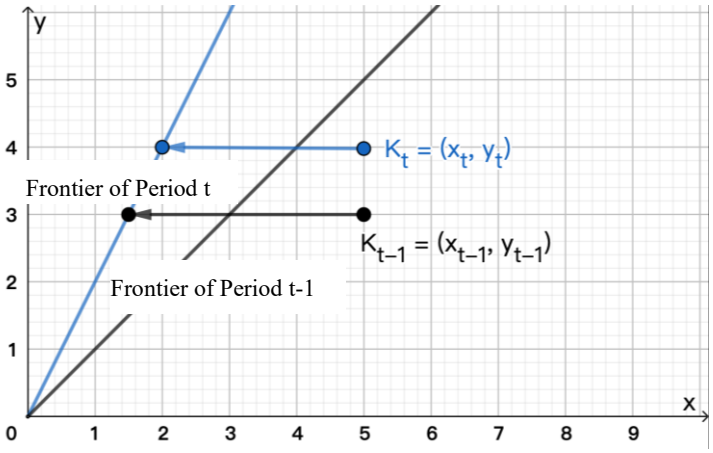


Figure 8-10 Schematic Diagram of the Malmquist Index (with the Frontier of Period t as Reference, CRS)

R Färe et al. (1992) followed the method ofCaves, Christensen, & Diewert (1982) in computing the Malmquist index, and adopted the geometric mean of the above two Malmquist indices as the Malmquist index of the evaluated DMU, that is,

$$MI(t\text{-}1,t)=\sqrt{MI^{t\text{-}1}_{(t\text{-}1,t)} \times MI^{t}_{(t\text{-}1,t)}}$$

$$=\sqrt{\frac{Score_t\text{-}1(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \times \frac{Score_t(x_t, y_t)}{Score_t(x_t\text{-}1, y_t\text{-}1)}} \quad (8.7)$$

$Score_t(x_t, y_t)$ and $Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)$ are respectively the technical

efficiency values of K in periods t and t-1, and their ratio is the technical efficiency change between the two periods (technical efficiency change, EC)

$$EC(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \quad (8.8)$$

The Malmquist index can be decomposed into technical efficiency change and technological change (TC), that is, MI = EC * TC.

$$MI(t,t\text{-}1)=\sqrt{\frac{Score_t\text{-}1(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t(x_t, y_t)}{Score_t(x_t\text{-}1, y_t\text{-}1)}}$$

$$=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\times\sqrt{\frac{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}{Score_t(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t\text{-}1(x_t, y_t)}{Score_t(x_t, y_t)}}$$

$$=EC(t\text{-}1,t)\times TC(t\text{-}1,t) \quad (8.9)$$

$$TC(t\text{-}1,t)=\sqrt{\frac{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}{Score_t(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t\text{-}1(x_t, y_t)}{Score_t(x_t, y_t)}} \quad (8.10)$$

In the output of MaxDEA, whether the model is input-oriented or output-oriented, the Malmquist index is interpreted in the same way: a value greater than 1 indicates an increase in productivity and a value smaller than 1 a decrease; the same applies to efficiency change (EC) and technological change (TC).

It should be noted that in practice the evaluated DMUs often come from a relatively small sample. Even though in theory the production technology being evaluated improves continuously over time, the frontier of the sample does not necessarily shift outward, and the technological change obtained from the DEA model is not necessarily technological progress.

8.4 Methods for Computing the Various Types of Malmquist Index

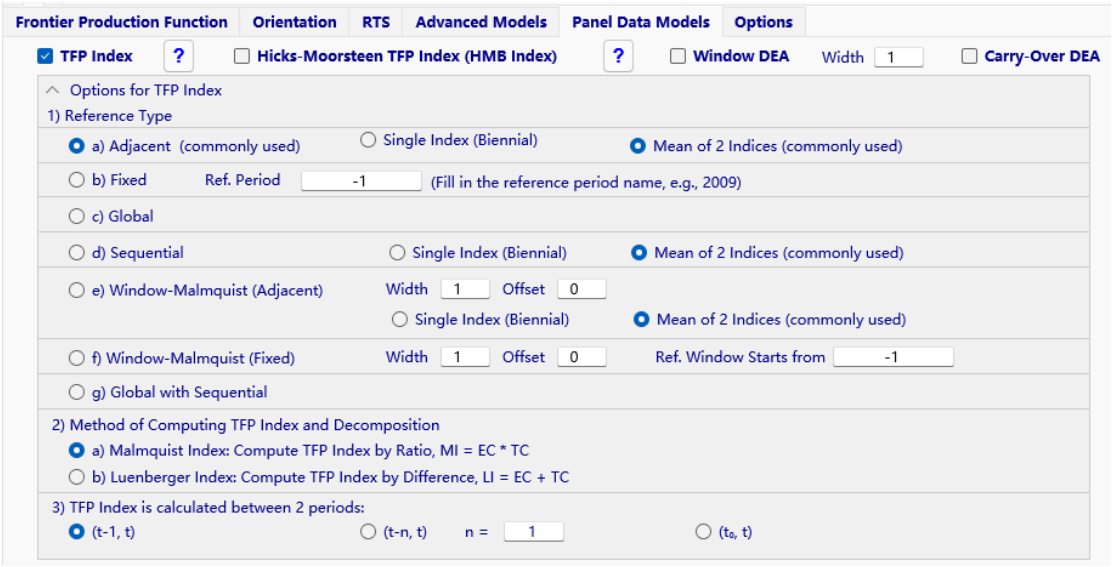


Figure 8-11 Types of MI in MaxDEA

MaxDEA provides 9 types of Malmquist index in 6 categories, which differ in the reference frontier. According to the way the reference frontier is determined, the Malmquist indices fall into two broad categories.

8.4.1 MI for Different Periods Computed against the Frontiers of Different Periods

The computation method described above references the frontiers of two adjacent periods (periods t and t-1) separately, obtains two TFP indices, and then takes their geometric mean as the Malmquist index. For example, MI(2019, 2020) uses the frontiers of 2019 and 2020, whereas MI(2018, 2019) uses the frontiers of 2018 and 2019. Three categories of Malmquist index use this rolling form of reference:

a) Adjacent benchmarking (most commonly used)

d) Sequential reference

e) Window benchmarking (adjacent)

8.4.1.1 Adjacent Reference (Average of Two Indices)

Adjacent reference is the classic MI computation method most commonly used in the literature, and the method described above is of this type.

The DEA efficiency values involved in the adjacent-reference MI(t-1, t) and their meanings are as follows. Let (x_t, y_t) denote the input and output values of period t, the corresponding

input and output values of period t-1 being (x_t-1, y_t-1); let Score_t() denote computation against the frontier of period t, and correspondingly Score_t-1() denote computation against the frontier of period t-1.

Score_t-1(x_t, y_t): the DEA efficiency value of (x_t, y_t) obtained against the frontier of period t-1;

Score_t-1(x_t-1, y_t-1): the DEA efficiency value of (x_t-1, y_t-1) obtained against the frontier of period t-1;

Score_t(x_t, y_t): the DEA efficiency value of (x_t, y_t) obtained against the frontier of period t;

Score_t(x_t-1, y_t-1): the DEA efficiency value of (x_t-1, y_t-1) obtained against the frontier of period t.

Table 8-4 Selected Output of the Malmquist Model (1)

t(Period)	DMU	Score_t(x_t, y_t)	Score_t-1(x_t-1, y_t-1)	Score_t-1(x_t, y_t)	Score_t(x_t-1, y_t-1)	MI(t-1, t)	EC(t-1, t)	TC(t-1, t)
2004	**P01**	**0.854874**	**0.726508**	**0.801142**	**0.775555**	**1.1025**	**1.1767**	**0.9370**
2004	P02	0.686738	0.609718	0.708859	0.590691	1.1626	1.1263	1.0322
2004	P03	1.000000	0.927464	1.015929	0.933164	1.0834	1.0782	1.0049
2004	P04	0.672176	0.644321	0.682884	0.632494	1.0613	1.0432	1.0173
2004	P05	1.000000	1.000000	1.095794	0.968794	1.0635	1.0000	1.0635
2004	P06	0.911377	0.808021	0.866099	0.839481	1.0787	1.1279	0.9564
2004	P07	0.823602	0.824668	0.836721	0.811737	1.0146	0.9987	1.0159
2004	P08	0.766695	0.675826	0.718719	0.718392	1.0654	1.1345	0.9391
2004	P09	0.854445	0.719902	0.800978	0.767957	1.1126	1.1869	0.9374
2004	P10	0.819708	0.717367	0.768415	0.765253	1.0712	1.1427	0.9374

Suppose the panel data cover 9 years, from 2003 to 2011. MI(t-1, t) is the ratio of the TFP of period t to the TFP of period t-1, and since the first year of the data is 2003, the earliest MI is MI(2003, 2004).

The computation of MI(2003, 2004) can be understood in conjunction with the output of MaxDEA. The column "t(Period)" in the results indicates which year's MI the current row computes, that is, which year t in MI(t-1, t) refers to. When t=2004, MI(t-1, t) = MI(2003, 2004), that is, the first row of the table.

$$\text{MI(t-1,t)}=\sqrt{\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\times\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t-1, y_t-1})}}=\sqrt{\frac{0.801142}{0.726508}\times\frac{0.854874}{0.775555}}=1.1025$$

$$\text{EC(t-1,t)}=\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}=\frac{0.854874}{0.726508}=1.1767$$

$$\text{TC(t-1,t)}=\sqrt{\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\times\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}}=\sqrt{\frac{0.726508}{0.775555}\times\frac{0.801142}{0.854874}}=0.9370$$

MaxDEA not only provides the TFP index (MI) and the various related DEA efficiency values involved in its computation, but also provides detailed results such as the projected values of these DEA models[i].

In the Productivity Index table, the detailed DEA model results corresponding to the column Score_t-1(x_t-1, y_t-1) and the column Score_t(x_t, y_t) are all contained in "Table Score_t(x_t, y_t)".

Score_t-1(x_t-1, y_t-1), t=2004 and Score_t(x_t, y_t), t=2003 denote the same DEA efficiency value; Score_t-1(x_t-1, y_t-1), t=2005 and Score_t(x_t, y_t), t=2004 denote the same DEA efficiency value.

8.4.1.2 Sequential Reference (Average of Two Indices)

The sequential reference Malmquist model was proposed byShestalova (2003) as a method of computing MI.

The formulas for sequential reference and for adjacent reference are exactly the same and are not repeated here. The difference between the two lies in the construction of the frontier. In adjacent reference, the frontier of period t is constructed from (x_t, y_t) and the frontier of period t-1 from (x_t-1, y_t-1), that is, only from the data of the current period; in sequential reference, the sequential frontier of period t is constructed from (x_t, y_t), (x_t-1, y_t-1), (x_t-2, y_t-2),..., (x_2, y_2), (x_1, y_1), and the sequential frontier of period t-1 from (x_t-1, y_t-1), (x_t-2, y_t-2),..., (x_2, y_2), (x_1, y_1), that is, the sequential frontier is constructed jointly from the data of the current period and of all preceding periods.

The way the sequential frontier is constructed means that the sequential frontier of period t envelops that of period t-1; hence the sequential frontier only moves forward with time and never regresses. This is an important feature of the sequential Malmquist model.

As the reference set expands, the efficiency values obtained from DEA models using radial distance, SBM (ERM), the directional distance function, MinDW or hybrid distance remain unchanged or decrease. The technological change (TC) obtained from the sequential Malmquist model is therefore never smaller than 1, that is, it always indicates technological progress.

Although the sequential frontier never regresses, if the MinDS distance is used in the model, the computed TC may still indicate technological regress. As explained above, a characteristic of the MinDS model is that the efficiency value may increase when the reference set expands (all

[i]Detailed results such as projected values are provided only in the envelopment model.

three cases, namely decrease, no change and increase, occur), which is not shared by other types of distance function. Hence, when the MinDS distance is selected, the technological change (TC) obtained from the sequential MI model may be smaller than 1 (all three cases, namely smaller than 1, equal to 1 and greater than 1, occur), that is, technological regress may still appear. Therefore, if technological regress is not wanted in the results, the MinDS distance type should not be used in the sequential MI model.

The construction of the sequential frontier is now illustrated with three periods as an example. To make the construction principle easy to demonstrate, a single-input single-output VRS model is used here. Figure 8-12 shows the current frontier of each of the three periods (periods 1, 2 and 3). The current frontier is constructed from the data of the current period: the period 1 frontier is A1B1C1D1E1; the period 2 frontier is A2C2D2E2; the period 3 frontier is A3B3D3E3.

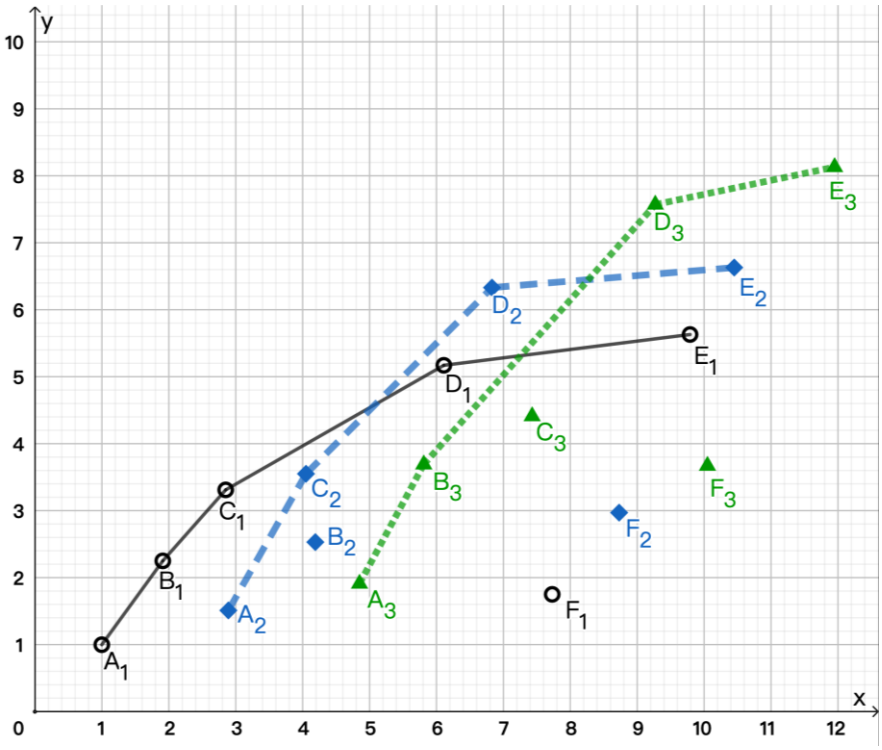


Figure 8-12 Current Frontiers of the Three Periods

Period 1 is the first period, so its sequential frontier is its current frontier. Figure 8-13 shows the sequential frontier of period 2,
A1B1C1D2E2, which is constructed jointly from the data of periods 1 and 2.

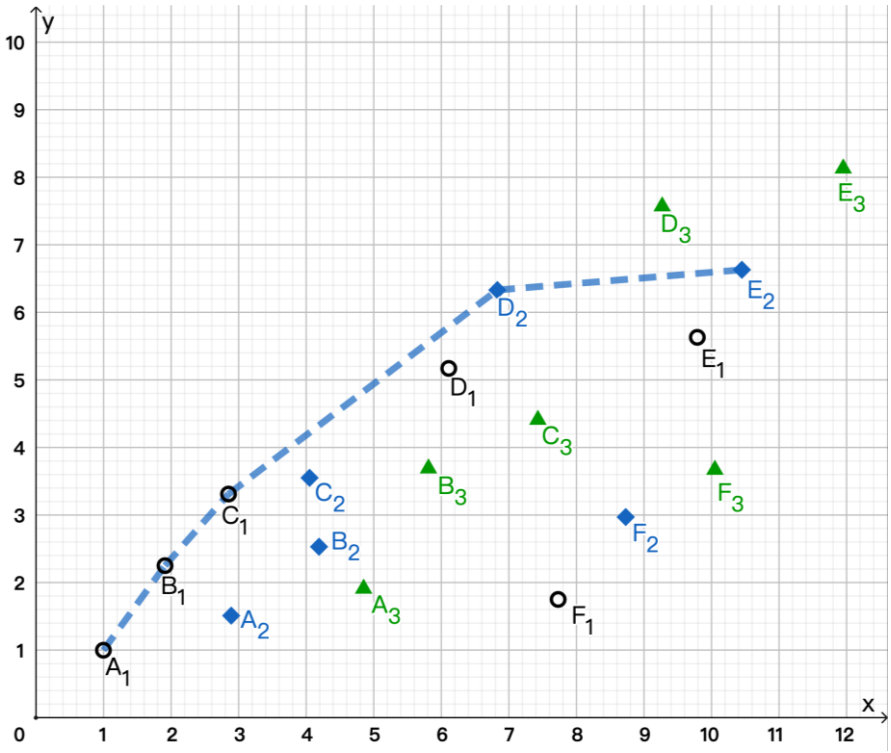


Figure 8-13 Sequential Frontier of Period 2

Figure 8-14 shows the sequential frontier of period 3, A1B1C1D2D3E3, which is constructed jointly from the data of periods 1, 2 and 3.

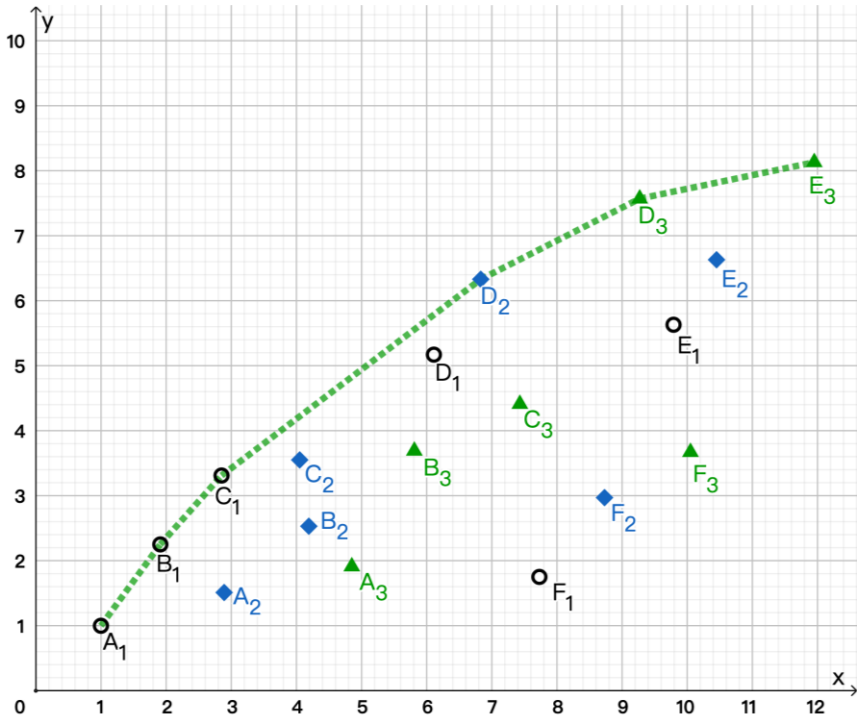


Figure 8-14 Sequential Frontier of Period 3

Figure 8-15 compares the sequential frontiers of the three periods and shows that the

sequential frontier always advances.

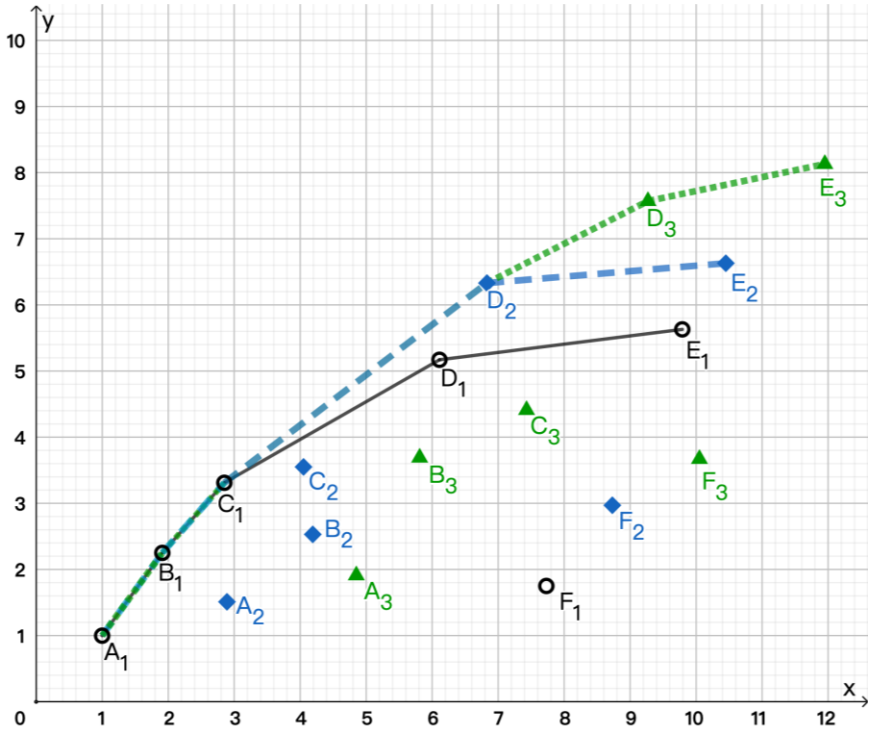


Figure 8-15 Comparison of the Sequential Frontiers of the Three Periods

8.4.1.3 Window Reference (Average of Two Indices)

Window benchmarking (adjacent) and adjacent benchmarking differ likewise in the construction of the frontier. The window frontier of period t is constructed jointly from the data of period t and of several preceding periods, and the window frontier of period t-1 is constructed jointly from the data of period t-1 and of several preceding periods(Thore, Kozmetsky, & Phillips, 1994; Asmild, Paradi, Aggarwall, & Schaffnit, 2004). The number of periods spanned by the window frontier (including the current period) is called the window width; for example, when the window comprises the current period and the two preceding periods, the window width is 3.

By default the window frontier comprises the current period and several preceding periods. If future periods are to be included, this can be achieved by means of the window offset. The offset denotes the number of periods by which the window is shifted towards the future. Taking a window width of 3 as an example, the 2010 window frontier (default offset = 0) comprises 2010, 2009 and 2008; if the offset is 1, the 2010 window frontier comprises 2011, 2010 and 2009; if the offset is 2, the 2010 window frontier comprises 2012, 2011 and 2010.

Taking a window width of 3 as an example, if no window offset is set (offset = 0), the window of each period contains the current period and the two preceding periods, so that the window width of the first two periods falls short of the value of 3 that was set. The actual window

width of period 1 is 1, containing only the current period; the actual window width of period 2 is 2; and the window width of period 3 and all later periods is 3. If the window offset is set to 1, the window of each period contains the current period together with one period before and one period after it. The actual window width of period 1 and of the last period is 2; the window width of the other periods (from period 2 to the penultimate period) is 3.

If the window width is set to 1, the window frontier of period t is the same as the contemporaneous frontier of period t, and window benchmarking is the same as the adjacent benchmarking introduced above.

The window of each period does not necessarily reach the set width. For example, when offset = 0 and the window width is set to 3, the window width of period 1 = 1[i] and the window width of period 2 = 2, and only period 3 and later periods reach a window width of 3.

If offset = 0 and the window width is set to its maximum (the total number of periods), the window frontier of period t is the same as the sequential frontier of period t, and window benchmarking is the same as sequential benchmarking.

All three methods of computing the MI introduced above involve cross-period benchmarking. Of the four DEA efficiency scores required to compute the MI, two are cross-period benchmarks (data of period t-1 benchmarked against the period t frontier, or data of period t benchmarked against the period t-1 frontier), namely $\mathrm{Score_t}(\mathrm{x_t\text{-}1}, \mathrm{y_t\text{-}1})$ and $\mathrm{Score_t\text{-}1}(\mathrm{x_t}, \mathrm{y_t})$. In these two DEA models the DMU under evaluation is not included in the reference set, so they are "naturally" super-efficiency models.

As Figure 8-16 shows, when the DMU under evaluation Kt= (xt, yt) is benchmarked against the frontier of period t-1, it lies outside the reference frontier and thus falls into the category of super-efficiency models.

8.4.1.4 Further Decomposition of TC

The method for the further decomposition of TC wasRolf Färe, Grifell-Tatjé, Grosskopf, & Knox Lovell (1997).

$$\mathrm{TC(t\text{-}1,t)}=\sqrt{\frac{\mathrm{Score_t\text{-}1}(\mathrm{x_t\text{-}1}, \mathrm{y_t\text{-}1})}{\mathrm{Score_t}(\mathrm{x_t\text{-}1}, \mathrm{y_t\text{-}1})}\times\frac{\mathrm{Score_t\text{-}1}(\mathrm{x_t}, \mathrm{y_t})}{\mathrm{Score_t}(\mathrm{x_t}, \mathrm{y_t})}}$$

[i]When the offset is 0, the window width of the starting period is always 1, however large the window width is set.

$$= \frac{\text{Score_t-1}(x_t\text{-}1, y_t\text{-}1)}{\text{Score_t}(x_t\text{-}1, y_t\text{-}1)} \times \sqrt{\frac{\left(\frac{\text{Score_t-1}(x_t, y_t)}{\text{Score_t}(x_t, y_t)}\right)}{\left(\frac{\text{Score_t-1}(x_t\text{-}1, y_t\text{-}1)}{\text{Score_t}(x_t\text{-}1, y_t\text{-}1)}\right)}}$$

$$= \text{MATC}(t\text{-}1, t) \times \text{BTC}(t\text{-}1, t) \qquad (8.11)$$

MATC(t-1, t) represents the magnitude of technological change: a value greater than 1 indicates technological progress (the frontier shifting outward), and a value less than 1 indicates technological regress (the frontier shifting inward). The TC computed with the data of period t-1, (x_t-1, y_t-1), measures the change in productivity brought about by a shift of the frontier on the assumption that the inputs and outputs of the DMU under evaluation remain unchanged. Compared with TC, MATC eliminates the bias in the measurement of technological change that arises from the changes in the input and output values between period t-1 and period t.

The relationship between MATC and TC is as follows: MATC can be understood simply as pure technological change, whereas TC contains confounding factors.

BTC (biased technological change) reflects the bias in the measurement of technological change that arises from the changes in the input and output values in period t.

BTC can be further decomposed into OBTC and IBTC.

$$BTC = \sqrt{\frac{\left(\frac{\text{Score_t-1}(x_t, y_t)}{\text{Score_t}(x_t, y_t)}\right)}{\left(\frac{\text{Score_t-1}(x_t\text{-}1, y_t\text{-}1)}{\text{Score_t}(x_t\text{-}1, y_t\text{-}1)}\right)}}$$

$$= \sqrt{\frac{\left(\frac{\text{Score_t-1}(x_t, y_t)}{\text{Score_t}(x_t, y_t)}\right)}{\left(\frac{\text{Score_t-1}(x_t, y_t\text{-}1)}{\text{Score_t}(x_t, y_t\text{-}1)}\right)}} \times \sqrt{\frac{\left(\frac{\text{Score_t-1}(x_t, y_t\text{-}1)}{\text{Score_t}(x_t, y_t\text{-}1)}\right)}{\left(\frac{\text{Score_t-1}(x_t\text{-}1, y_t\text{-}1)}{\text{Score_t}(x_t\text{-}1, y_t\text{-}1)}\right)}}$$

$$= \text{OBTC}(t\text{-}1,t) \times \text{IBTC}(t\text{-}1,t) \qquad (8.12)$$

where Score_t-1(x_t, y_t-1) denotes the DEA efficiency score of the virtual DMU with mixed data [xt, yt-1] benchmarked against the frontier of period t-1 (with the data of period t-1 as the reference set), and Score_t(x_t, y_t-1) denotes the DEA efficiency score of the virtual DMU with mixed data [xt, yt-1] benchmarked against the frontier of period t (with the data of period t as the reference set).

OBTC(t-1, t) denotes output biased technological change, meaning the bias in the measurement of technological change caused, in period t relative to period t-1, by the change in

output (that is, by y_t-1 becoming y_t).

IBTC(t-1, t) denotes input biased technological change, meaning the bias in the measurement of technological change caused, in period t relative to period t-1, by the change in input (that is, by x_t-1 becoming x_t).

8.4.1.5 The Infeasibility Problem of the VRS Model Caused by Cross-Period Benchmarking[i]

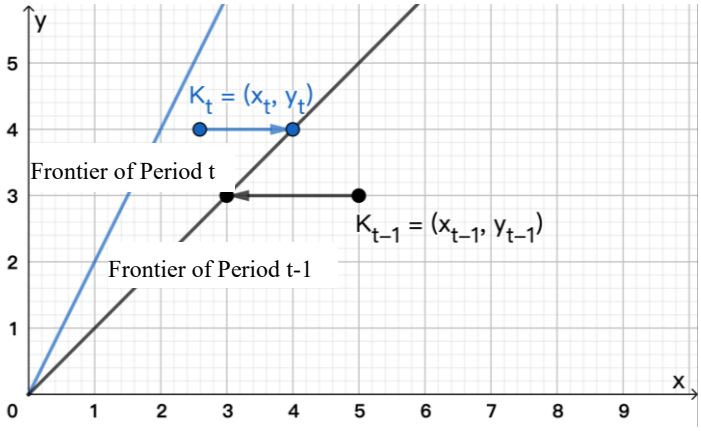


Figure 8-16 Super-efficiency Illustration in the Computation of the Malmquist Index (Benchmarked against the Period t-1 Frontier, CRS)

As introduced in Chapter 5, input- or output-oriented VRS super-efficiency models suffer from the infeasibility problem; likewise, the VRS DEA models involved in the Malmquist model suffer from the infeasibility problem, that is, in the VRS model the computation of $\text{Score_t}(\text{x_t-1}, \text{y_t-1})$ and $\text{Score_t-1}(\text{x_t}, \text{y_t})$ may be infeasible. For the infeasibility problem, one option is to adopt the solutions introduced in Chapter 5 for super-efficiency models, and the other is to adopt the biennial frontier benchmarking (single TFP index) method introduced below.

8.4.1.6 Biennial Frontier (Single Index)

Biennial benchmarking uses the frontier constructed jointly from the data of two adjacent periods, (x_t-1,y_t-1) and (x_t,y_t), as the reference frontier (Jesús T. Pastor, Asmild, & Lovell, 2011).(Jesús T. Pastor, Asmild, & Lovell, 2011) Because the biennial frontier envelops all the DMUs under evaluation in the two periods, the situation described above, in which a DMU under evaluation lies outside the frontier, does not arise, and the possible infeasibility of the VRS model

[i]A DEA model may be infeasible for various reasons; here the reference is specifically to the infeasibility of the VRS model when the DMU under evaluation lies beyond the frontier.

is avoided.

Let Score_t-1Ut() denote the DEA efficiency score obtained against the biennial frontier; Score_t-1Ut(x_t, y_t) denotes the DEA efficiency score of (x_t, y_t) obtained against the biennial frontier; and Score_t-1Ut(x_t-1, y_t-1) denotes the DEA efficiency score of (x_t-1, y_t-1) obtained against the biennial frontier.

The data of two adjacent periods are benchmarked against the same biennial frontier, so there is only one TFP index and it is no longer necessary to compute two productivity indices and take their geometric mean. The biennial MI(t-1,t) is

$$\mathrm{MI(t\text{-}1,t)}=\frac{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)} \quad (8.13)$$

EC(t-1,t) is computed as before. Although MI(t-1,t) for two adjacent periods is computed against a single biennial frontier, each of the two periods still has its own frontier (contemporaneous, sequential or window frontier), and EC(t-1,t) is still computed by taking the technical efficiency scores of each period against its contemporaneous frontier and then forming their ratio.

$$\mathrm{EC(t\text{-}1,t)}=\frac{\mathrm{Score_t}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t\text{-}1}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)} \quad (8.14)$$

The MI can still be decomposed into EC and TC.

$$\mathrm{TC(t\text{-}1,t)}=\frac{\mathrm{MI(t\text{-}1,t)}}{\mathrm{EC(t\text{-}1,t)}}$$

$$=\frac{\left(\frac{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)}\right)}{\left(\frac{\mathrm{Score_t}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t\text{-}1}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)}\right)}$$

$$=\frac{\left(\frac{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t}\left(\mathrm{x_t, y_t}\right)}\right)}{\left(\frac{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)}{\mathrm{Score_t\text{-}1}\left(\mathrm{x_t\text{-}1, y_t\text{-}1}\right)}\right)} \quad (8.15)$$

where $\frac{\mathrm{Score_t\text{-}1Ut}\left(\mathrm{x_t, y_t}\right)}{\mathrm{Score_t}\left(\mathrm{x_t, y_t}\right)}$ reflects the distance between the period t frontier and the biennial frontier, a larger ratio indicating that the period t frontier is closer to the biennial

frontier; $\frac{\text{Score_t-1Ut}(\text{x_t-1, y_t-1})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}$ reflects the distance between the period t-1 frontier and the biennial frontier, a larger ratio indicating that the period t-1 frontier is closer to the biennial frontier. The ratio of the two then reflects the change in the period t frontier relative to the period t-1 frontier.

8.4.1.7 Further Decomposition of TC under the Biennial Frontier

$$\text{TC(t-1,t)} = \frac{\left(\frac{\text{Score_t-1Ut}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1Ut}(\text{x_t-1, y_t-1})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\right)}$$

$$= \left(\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\right) \times \frac{\left(\frac{\text{Score_t-1Ut}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1Ut}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\right)}$$

$$= MATC \times BTC$$

BTC (biased technological change) reflects the bias that arises, when the distance between the period t frontier and the biennial frontier is measured, from the use of indirect rather than direct measurement and from the different observation points (x_t-1, y_t-1 and x_t, y_t). BTC can also be further decomposed into OBTC and IBTC by the method given above, but this is of little value and there is no need for further decomposition.

When the sequential frontier method is chosen, because the sequential frontier of period t envelops the sequential frontier of period t-1, the biennial sequential frontier of two adjacent periods is identical to the sequential frontier of period t, that is, for the sequential frontier Score_t-1Ut() = Score_t().

The computation of TC under the biennial sequential frontier then simplifies to

$$\text{TC(t-1,t)} = \frac{\left(\frac{\text{Score_t-1Ut}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1Ut}(\text{x_t-1, y_t-1})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\right)}$$

$$= \frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})} = MATC$$

Therefore, for the biennial sequential frontier, TC = MATC and BTC = 1, and no further decomposition is required. The biennial sequential frontier is the only benchmarking method that

produces no bias in the measurement of TC.

8.4.1.8 The Non-multiplicability Problem of the MI

The Malmquist index with adjacent benchmarking is not circular: an MI(t-n,t) spanning several periods (n>1) cannot be computed by multiplying the MI(t-1,t) of the successive periods, for example

$MI(t-2,t) \neq MI(t-2,t-1) \times MI(t-1,t)$.

If the panel data contain 10 years of annual data from 2011 to 2020, the MI computation methods introduced above yield MI(t-1,t) for the 9 years from 2012 to 2020 relative to the preceding year. However, the product of the MIs of these 9 years is not equal to MI(2011,2020), the index comparing 2020 with 2011. The following introduces MI computation methods that can be multiplied cumulatively, that is, MI(t-1,t) for all periods is computed against the same frontier.

8.4.2 Computing the MI of All Periods against the Same Frontier

When the Malmquist index is computed against the same frontier, there are three types of Malmquist index:

b) Fixed benchmarking

c) Global benchmarking

f) Window benchmarking (fixed)

A fixed frontier is the frontier of a certain fixed period; a global frontier is the frontier constructed jointly from all periods; and a window frontier (fixed) is the frontier of a fixed window period (a time window consisting of several consecutive periods).

The MI, EC and TC obtained from these three benchmarking methods are circular and can be multiplied cumulatively, that is

$MI(t-n,t) = MI(t-n, t-n+1) \times MI(t-n+1, t-n+2) \times \cdots \times MI(t-2,t) \times MI(t-1,t)$.

8.4.2.1 Fixed Benchmarking (Single Index)

The fixed reference Malmquist model was proposed byBerg, Forsund, & Jansen (1992) and is an MI computation method that uses the contemporaneous frontier of a certain fixed period as the reference frontier for computing MI(t-1,t) in every period.

All periods are benchmarked against the same fixed frontier, so there is only one TFP index. The fixed-benchmark MI(t-1,t) is

$$MI(t\text{-}1,t)=\frac{Score_f(x_t, y_t)}{Score_f(x_t\text{-}1, y_t\text{-}1)} \quad (8.16)$$

Score_f (x_t, y_t) denotes the DEA efficiency score of (x_t, y_t) obtained against the fixed frontier.

Score_f (x_t-1, y_t-1) denotes the DEA efficiency score of (x_t-1, y_t-1) obtained against the fixed frontier.

EC(t-1,t) is still computed as follows:

$$EC(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \quad (8.17)$$

The MI can still be decomposed into EC and TC.

$$TC(t\text{-}1,t)=\frac{MI(t\text{-}1,t)}{EC(t\text{-}1,t)}$$

$$=\frac{\left(\frac{Score_f(x_t, y_t)}{Score_f(x_t\text{-}1, y_t\text{-}1)}\right)}{\left(\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\right)}$$

$$=\frac{\left(\frac{Score_f(x_t, y_t)}{Score_t(x_t, y_t)}\right)}{\left(\frac{Score_f(x_t\text{-}1, y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\right)} \quad (8.18)$$

where $\frac{Score_f(x_t, y_t)}{Score_t(x_t, y_t)}$ reflects the distance between the period t frontier and the fixed frontier, a larger ratio indicating that the period t frontier is closer to the fixed frontier; $\frac{Score_f(x_t\text{-}1, y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}$ reflects the distance between the period t-1 frontier and the fixed frontier, a larger ratio indicating that the period t-1 frontier is closer to the fixed frontier. The ratio of the two then reflects the change in the period t frontier relative to the period t-1 frontier.

8.4.2.2 Global Benchmarking (Single Index)

The global reference Malmquist model was proposed byJesús T. Pastor & Lovell (2005) and is a method of computing the Malmquist index that uses the frontier constructed jointly from all periods (the global frontier) as the reference frontier.

All periods are benchmarked against the same global frontier, so there is only one TFP index. The global-benchmark MI(t-1,t) is

$$\text{MI(t-1,t)}=\frac{\text{Score_g}(\text{x_t, y_t})}{\text{Score_g}(\text{x_t-1, y_t-1})} \quad (8.19)$$

Score_g (x_t, y_t) denotes the DEA efficiency score of (x_t, y_t) obtained against the global frontier.

Score_g (x_t-1, y_t-1) denotes the DEA efficiency score of (x_t-1, y_t-1) obtained against the global frontier.

EC(t-1,t) is still computed as follows:

$$\text{EC(t-1,t)}=\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})} \quad (8.20)$$

The MI can still be decomposed into EC and TC.

$$\text{TC(t-1,t)}=\frac{\text{MI(t-1,t)}}{\text{EC(t-1,t)}}$$

$$=\frac{\left(\frac{\text{Score_g}(\text{x_t, y_t})}{\text{Score_g}(\text{x_t-1, y_t-1})}\right)}{\left(\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\right)}$$

$$=\frac{\left(\frac{\text{Score_g}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_g}(\text{x_t-1, y_t-1})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\right)} \quad (8.21)$$

where $\frac{\text{Score_g}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}$ reflects the distance between the period t frontier and the global frontier, a larger ratio indicating that the period t frontier is closer to the global frontier; $\frac{\text{Score_g}(\text{x_t-1, y_t-1})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}$ reflects the distance between the period t-1 frontier and the global frontier, a larger ratio indicating that the period t-1 frontier is closer to the global frontier. The ratio of the two then reflects the change in the period t frontier relative to the period t-1 frontier.

Since the "DMU" under evaluation is necessarily contained in the global reference set, the global reference Malmquist index does not suffer from the infeasibility problem of the VRS

model.

8.4.2.3 Fixed Window Benchmarking

"f) Window benchmarking (fixed)" requires three parameters, of which the window width and the offset are the same as in "e) Window benchmarking (adjacent)". The other parameter sets the starting period of the fixed window; the ending period is inferred automatically by the software from the window width.

The computation of the MI under the fixed window frontier is similar to that under the fixed frontier described above, with two differences:

1) In computing MI(t-1,t), the fixed frontier is the frontier constructed for a certain fixed period, whereas the fixed window frontier is the frontier constructed for a fixed window period.

2) In computing EC(t-1,t), fixed benchmarking involves the contemporaneous frontiers of two adjacent periods, whereas fixed window benchmarking involves the window frontiers of two adjacent periods.

8.4.2.4 Further Decomposition of TC under the Same Frontier

Further decomposition of TC under global benchmarking:

$$\mathrm{TC(t\text{-}1,t)}=\frac{\left(\frac{\mathrm{Score_g(x_t,\,y_t)}}{\mathrm{Score_t(x_t,\,y_t)}}\right)}{\left(\frac{\mathrm{Score_g(x_t\text{-}1,\,y_t\text{-}1)}}{\mathrm{Score_t\text{-}1(x_t\text{-}1,\,y_t\text{-}1)}}\right)}$$

$$=\left(\frac{\mathrm{Score_t\text{-}1(x_t\text{-}1,\,y_t\text{-}1)}}{\mathrm{Score_t(x_t\text{-}1,\,y_t\text{-}1)}}\right)\times\frac{\left(\frac{\mathrm{Score_g(x_t,\,y_t)}}{\mathrm{Score_t(x_t,\,y_t)}}\right)}{\left(\frac{\mathrm{Score_g(x_t\text{-}1,\,y_t\text{-}1)}}{\mathrm{Score_t(x_t\text{-}1,\,y_t\text{-}1)}}\right)}$$

$$=MATC\mathrm{(t\text{-}1,t)}\times BTC\mathrm{(t\text{-}1,t)}$$

Further decomposition of TC under fixed benchmarking and fixed window benchmarking:

$$\mathrm{TC(t\text{-}1,t)}=\frac{\left(\frac{\mathrm{Score_f(x_t,\,y_t)}}{\mathrm{Score_t(x_t,\,y_t)}}\right)}{\left(\frac{\mathrm{Score_f(x_t\text{-}1,\,y_t\text{-}1)}}{\mathrm{Score_t\text{-}1(x_t\text{-}1,\,y_t\text{-}1)}}\right)}$$

$$= \left(\frac{\text{Score_t-1}\left(\text{x_t-1, y_t-1}\right)}{\text{Score_t}\left(\text{x_t-1, y_t-1}\right)} \right) \times \frac{\left(\frac{\text{Score_f}\left(\text{x_t, y_t}\right)}{\text{Score_t}\left(\text{x_t, y_t}\right)} \right)}{\left(\frac{\text{Score_f}\left(\text{x_t-1, y_t-1}\right)}{\text{Score_t}\left(\text{x_t-1, y_t-1}\right)} \right)}$$

$$= MATC(\text{t-1,t}) \times BTC(\text{t-1,t})$$

Note that MATC and BTC are not circular and cannot be multiplied cumulatively.

8.5 Comparison of the Various MIs

1) Adjacent benchmarking (the average of two indices) is the earliest and most widely applied MI; although it suffers from the infeasibility problem of the VRS model, the methods for dealing with that problem introduced above can be used.[i]

2) The biennial frontier and global benchmarking MI models do not suffer from the infeasibility problem of the VRS model.

3) Window benchmarking (the average of two indices) also suffers from the infeasibility problem of the VRS model. However, when the window width is w (w⩾3) and the offset is set to a value between 1 and w-2 (for example, when the window width is 3 the offset can only be set to 1, and when the window width is 4 the offset can be set to 1 or 2), the infeasibility problem can be avoided.

4) The sequential benchmarking MI model avoids results that show technological regress (except for the MinDS distance type).

5) Global, sequential and window benchmarking increase the number of "DMUs" in the computation, so when the number of DMUs is insufficient these three types of model can be used to improve the discriminating power of the results. It should be noted that in the sequential MI model the increase in the number of "DMUs" is uneven across periods: there is no increase in period 1, the increase becomes larger in later periods, and it is largest in the last period, where the number of "DMUs" equals that in the global MI model. In empirical research the number of DMUs is often not sufficiently large, and the frontier established by the model is sometimes formed by only a few DMUs, so that the frontier is not refined enough and the results obtained are of poor reliability or stability. In such cases, MI models that increase the number of "DMUs" in this way should be chosen in order to refine the frontier and thereby improve the reliability and stability of the results.

6) The fixed benchmarking, global benchmarking and fixed window benchmarking MIs can be multiplied cumulatively. The cumulative product of MI(t-1,t) for period t and the MIs of all the preceding periods yields MI(0,t), that is, the MI of period t relative to period 0 (the base period or first period).

$$\prod_{t=1}^{T} MI\left(t-1,t\right)=MI\left(0,\mathrm{T}\right)$$

Taking global benchmarking as an example,

[i]The literature applying this type of Malmquist model generally does not mention the infeasibility problem; in practice the efficiency score is often set to 1 when the model is infeasible.

MI(0,1) × MI(1,2) × … × MI(t-2,t-1) × MI(t-1,t)

= Score_g(x_1,y_1) / Score_g(x_0,y_0) × Score_g(x_2,y_2) / Score_g(x_1,y_1) × … × Score_g(x_t-1,y_t-1) / Score_g(x_t-2,y_t-2) × Score_g(x_t,y_t) / Score_g(x_t-1,y_t-1)

= Score_g(x_t,y_t) / Score_g(x_0,y_0)

= MI(0,t)

7) The fixed-benchmark efficiency scores Score_f() or global-benchmark efficiency scores Score_g() of each period obtained from the three MI models, namely fixed benchmarking, global benchmarking and fixed window benchmarking, are all measured against the same frontier and are therefore comparable with one another in magnitude.[i] Although Score_f() and Score_g() are not themselves TFP, they can reflect the relative magnitude of TFP.

MI(t-1,t) is the ratio of the TFP of a DMU in period t to the TFP of the same DMU in period t-1, so it reflects only the rate of change in the TFP of that DMU between the two periods and cannot reflect the level of that DMU's TFP in either of the two periods. The cumulative product MI(0,t) is the ratio of the TFP of a DMU in period t to the TFP of the same DMU in period 0 (the base period or first period), so it reflects only the rate of change in the TFP of that DMU between period t and period 0 and cannot reflect the level of that DMU's TFP in period t or period 0. Therefore, the MI(0,t) of a DMU can be used only to compare the TFP of that DMU with itself over different periods, and cannot be used to compare the TFP of different DMUs in different periods.

If the researcher is interested in the TFP growth rate of each DMU in each year, the answer is MI(t-1,t); if the cumulative TFP growth rate of each DMU is also of interest, the answer is MI(0,t). If the researcher wishes to compare the level of TFP of different DMUs in different years, a different indicator is required. The fixed-benchmark efficiency score Score_f() or the global-benchmark efficiency score Score_g() can serve this purpose.

Taking global benchmarking as an example, the global-benchmark efficiency score Score_g(x_t,y_t) is essentially the ratio of the TFP of the DMU under evaluation to the TFP of the projection point of that DMU on the global frontier, and is therefore also essentially a TFP index, that is

$$\text{Score_g}(\text{x_t},\text{y_t})=\frac{\text{TFP}(x_t,y_t)}{\text{TFP}(\text{benchmark}_g)}=\frac{\text{TFP}(x_t,y_t)}{\text{TFP}(\hat{x}_g,\hat{y}_g)}.$$

$(\hat{x}_g,\hat{y}_g)$ is the projection point of the DMU under evaluation on the global frontier. Although each DMU under evaluation has a different projection point on the frontier, the TFP of

[i]The TFP value itself cannot be calculated by means of DEA.

the projection points of different DMUs on the production frontier is equivalent. It can therefore be considered that the denominator is the same when different DMUs compute the above ratio. Alternatively, Score_g(x_t,y_t) can be understood as the ratio of the TFP(x_t,y_t) of each DMU in each period to the maximum TFP among all the years in the sample, so the maximum value of this ratio is 1. Although Score_g(x_t,y_t) is not TFP, it can reflect the relative magnitude of TFP of different DMUs in different periods.

Line charts, bar charts and similar figures can be produced from the fixed-benchmark efficiency scores Score_f(x_t,y_t) or the global-benchmark efficiency scores Score_g(x_t,y_t) to observe the changes in TFP over periods intuitively, for example, Figure 8-17 is a comparison of Score_g(x_t,y_t) for hospitals in the eastern, central and western regions from 2003 to 2011 (the three curves in the figure are the averages of Score_g(x_t,y_t) of the provinces within the eastern, central and western regions, so the maximum value in the figure does not reach 1).

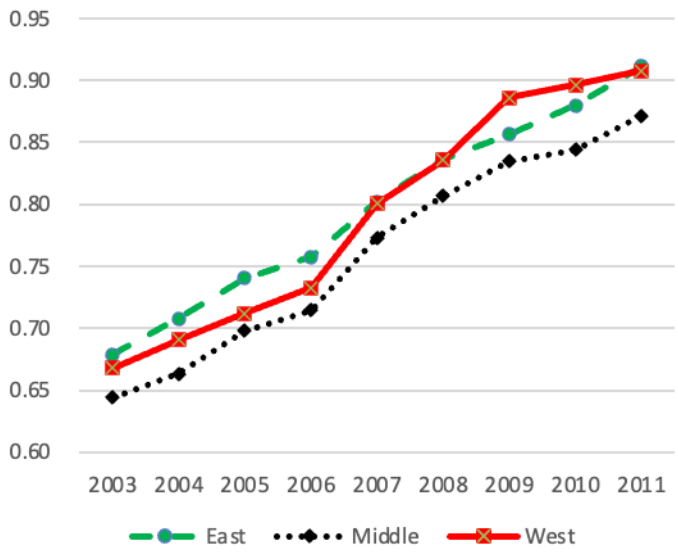


Figure 8-17 Score_g(x_t,y_t) of Hospitals in the Eastern, Central and Western Regions, 2003-2011

8) If the conclusions drawn from the analysis results of different MI models differ greatly, it is very likely that the frontier is too coarse because the number of DMUs is insufficient. Theoretically, as the number of DMUs increases (assuming that the data contain no measurement error), the conclusions drawn from the analysis results of different types of MI tend to coincide.

Table 8-5 Comparison of the Characteristics of the Various MIs

Benchmarking method	Index computation	MI(t-1,t) computation Frontier involved	Whether the MIs of all periods are benchmarked against the same frontier	Cumulative product MI(t-1,t) = MI(0,t)	EC(t-1,t) computation Frontier involved	Further decomposition of TC*	VRS infeasibility problem	Indicator usable for comparing TFP across DMUs and periods
a) Adjacent benchmarking	Average of two indices	Contemporaneous frontier	No	No	Contemporaneous frontier	**Yes**	Present	None
	Single index	Biennial frontier	No	No	Contemporaneous frontier	No	Absent	None
b) Fixed benchmarking	Single index	Fixed frontier	**Yes**	**Yes**	Contemporaneous frontier	No	Present	Score_f(x_t,y_t)
c) Global benchmarking	Single index	Global frontier	**Yes**	**Yes**	Contemporaneous frontier	No	Absent	Score_g(x_t,y_t)
d) Sequential benchmarking	Average of two indices	Sequential frontier	No	No	Sequential frontier	**Yes**	Present	None
	Single index	Biennial sequential frontier**	No	No	Sequential frontier	No	Absent	None
e) Window benchmarking	Average of two indices	Window frontier	No	No	Window frontier	**Yes**	Present, but avoidable by setting	None

(adjacent)							the window offset***	
	Single index	Biennial window frontier	No	No	Window frontier	No	Absent	None
f) Window benchmarking (fixed)	Single index	Fixed window frontier	**Yes**	**Yes**	Window frontier	No	Present	Score_f(x_t,y_t)

*** TC(t-1,t) = OBTC(t-1,t) * IBTC(t-1,t) * MATC(t-1,t). The frontiers involved in the computation of TC(t-1,t) include the frontier involved in the computation of MI(t-1,t) and the frontier involved in the computation of EC(t-1,t), and are not listed separately.**

**** The biennial sequential frontier of periods t-1 and t is identical to the sequential frontier of period t.**

***** When the window width is 3, setting the offset to 1 avoids the VRS infeasibility problem; when the window width is 4, setting the offset to 1 or 2 avoids the VRS infeasibility problem; and when the window width is w, setting the offset to a value between 1 and w-2 avoids the VRS infeasibility problem. A DEA model may be infeasible for various reasons; here the term refers specifically to the infeasibility of the VRS model that occurs when the DMU under evaluation lies outside the frontier.**

8.6 Malmquist Index Decomposition Methods Commonly Used in the Literature

The preceding sections have set out in detail the formulas for decomposing each type of MI into the product of EC and TC; this decomposition does not take scale efficiency into account. When scale efficiency is taken into account, three further MI decomposition methods are commonly used in the literature. On the returns-to-scale interface, if "6) Decomposition of efficiency or TFP indices (including scale efficiency/scale effect)" is selected, four MI decomposition results can be seen in the output interface of MaxDEA, each in a separate table.

FGLR: the decomposition method ofR Färe et al. (1992) described in detail above, which does not involve scale efficiency.

FGNZ:R. Färe, Grosskopf, Norris, & Zhang (1994) which further decomposes EC on the basis of FGLR.

RD:Ray & Desli (1997).

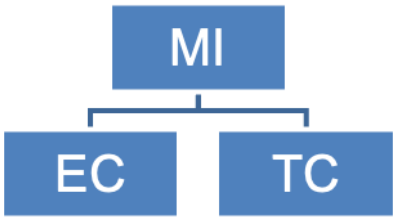
Zofio:Zofio (2007).

If the average MI of each period (the average over all DMUs in a period) or the average MI of each DMU (the average over all periods for a DMU) is required, geometric averaging must be used. Geometric averaging still preserves the decomposition MI = EC × TC.

Note that according to the definition of scale efficiency, a radial distance must be used to compute scale efficiency. When other distance types are used, the resulting TE/PTE ratio is not "scale efficiency"; in MaxDEA this analogous ratio is called "scale effect" to distinguish it (see Section 2.6 on the scale efficiency problem in Chapter 2). If the distance type used is not a radial distance, the SEC obtained in the decomposition methods below is not "scale efficiency change" but "scale effect change".

8.6.1 R Färe et al. (1992) Decomposition Method

R Färe et al. (1992) decompose the Malmquist index into technical efficiency change (EC) and technological change (TC).



This is the most basic decomposition method, and it does not involve scale efficiency. This decomposition method has already been described in detail in the preceding sections on the various MIs, namely

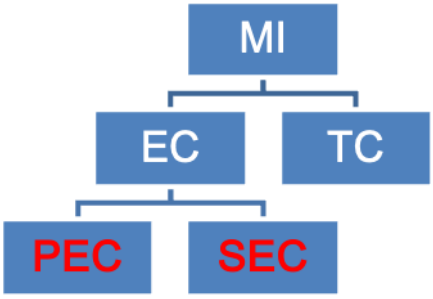
MI(t-1,t) = EC(t-1,t) * TC(t-1,t)

8.6.2 R. Färe et al. (1994) Decomposition Method

R. Färe et al. (1994), building on the decomposition method of R Färe et al. (1992), use the different efficiency change values (EC) obtained from the VRS Malmquist and CRS Malmquist models to further decompose itsR Färe et al. (1992) EC into pure technical efficiency change (PEC) and scale efficiency

change (SEC), namely

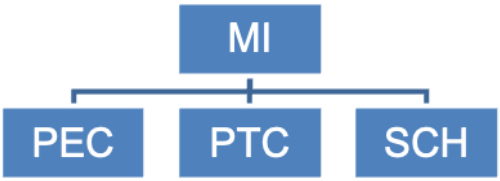
MI = EC * TC = PEC * SEC * TC



8.6.3 Ray & Desli (1997) Decomposition Method

Ray & Desli (1997) decompose the Malmquist index obtained from the CRS model into pure efficiency change (PEC), pure technological change (PTC) and a scale change factor, namely

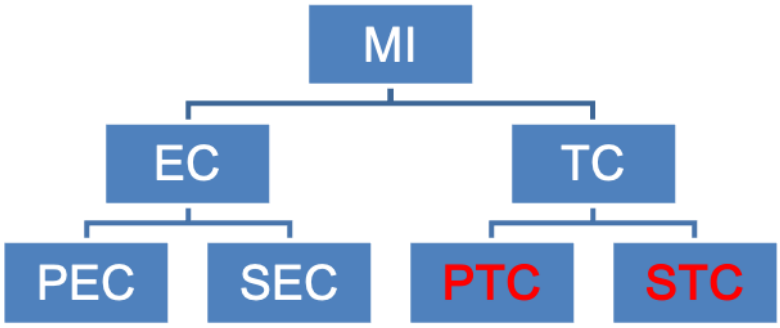
MI = PEC * PTC*SCH



8.6.4 Zofio (2007) Decomposition Method

R. Färe et al. (1994); Zofio (2007) the decomposition method of R. Färe et al. (1994), further decomposing TC into pure technological change (PTC) and scale technological change (STC), namely

MI = PEC * SEC * TC=PEC * SEC * PTC*STC



8.7 The Luenberger Index

Among the options of the TFP index model there is also an item "Method for Calculating and Decomposing the TFP Index", which offers two choices:

1) Malmquist index

2) Luenberger index

The TFP indices introduced above are all calculated by the ratio method and are called Malmquist indices (MI). The decomposition of MI into EC and TC is multiplicative. An MI greater than 1 indicates an increase in productivity, and an MI less than 1 indicates a decrease in productivity.

If the difference method is used to calculate the TFP index(Chambers, Färe, & Grosskopf, 1996), it is called the Luenberger index (LI), and the decomposition of LI into EC and TC is additive. An LI greater than 0 indicates an increase in productivity, and an LI less than 0 indicates a decrease in productivity.[i] For example, the formula for the adjacent benchmarking (average of two indices) Luenberger index is

$$LI(t\text{-}1,t)=\frac{\left(Score_t\text{-}1\left(x_t, y_t\right)-Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)+Score_t\left(x_t, y_t\right)-Score_t\left(x_t\text{-}1, y_t\text{-}1\right)\right)}{2}$$

$$EC(t\text{-}1,t)=Score_t\left(x_t, y_t\right)-Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)$$

$$TC(t\text{-}1,t)=\frac{\left(Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)-Score_t\left(x_t\text{-}1, y_t\text{-}1\right)+Score_t\text{-}1\left(x_t, y_t\right)-Score_t\left(x_t, y_t\right)\right)}{2}$$

$$LI(t\text{-}1,t)=EC(t\text{-}1,t)+TC(t\text{-}1,t)$$

Chambers et al. (1996) adopted the difference method because the productivity index in that paper is calculated using a directional distance function, with β^* used to compute the productivity index, and since β^* may be positive, negative or 0, the ratio method cannot be used to calculate a TFP index. This method is now rarely used. Unless it is essential, the Luenberger index is not recommended.

If the average LI for each period (the mean of all DMUs in that period) or the average LI for each DMU (the mean of one DMU over all periods) is required, an arithmetic mean must be used. The arithmetic mean still preserves the decomposition LI = EC + TC.

8.8 The Malmquist-Luenberger Index

Chung et al. (1997) applied the directional distance function incorporating undesirable outputs to the Malmquist model and called the resulting Malmquist index the Malmquist-Luenberger index (ML index for short). The Malmquist index obtained from any Malmquist model that includes undesirable outputs can be called a Malmquist-Luenberger index; that is, in the MaxDEA software, combining an undesirable output model with the Malmquist model yields the Malmquist-Luenberger index.Chung et al. (1997) used the directional distance function and the adjacent benchmarking (average of two indices) Malmquist model, whereas in the MaxDEA software any other distance function, such as radial, SBM (ERM) or MinDS, can be chosen for combination with the Malmquist model; other types of Malmquist index, such as global benchmarking, can also be selected.

Note that the ML index belongs to the Malmquist index, so "Malmquist index" must be selected under "Method for Calculating and Decomposing the TFP Index".

[i]Note that in panel data models, if a weighted additive distance is used, the results obtained with all models other than RAM and BAM are "inefficiency values", so that only the Luenberger index can be used; moreover, the Luenberger index obtained in this case has the opposite meaning, a value less than 0 indicating an increase in productivity and a value greater than 0 a decrease.

8.9 The Cost-Malmquist Index

Maniadakis & Thanassoulis (2004) combined the cost model with the Malmquist index to obtain the cost-Malmquist index. In the MaxDEA software, the revenue model and others can also be combined with the Malmquist index to obtain measures such as the revenue-Malmquist index, and other types of Malmquist index, such as global benchmarking, can also be selected.

The decomposition of the cost Malmquist model proceeds in two steps, and the MaxDEA software directly provides the results of the first step:

Step 1: Run the cost efficiency Malmquist model; the results contain

- Malmquist Index (the cost Malmquist index) = Efficiency Change (the cost efficiency change) * Technological Change (the overall technological change)

Step 2: Run the radial Malmquist model, and then use the results of the two steps to carry out the following decomposition:

- Allocative Efficiency Change (allocative efficiency change) = Efficiency Change (from the cost Malmquist model) / Efficiency Change (from the radial Malmquist model)
- Price Effect (price effect) = Technological Change (from the cost Malmquist model) / Technological Change (from the radial Malmquist model)

The overall decomposition of the cost Malmquist Index is

- Cost Malmquist Index = Efficiency Change (technical efficiency change) * Allocative Efficiency Change (allocative efficiency change) * Technological Change (technological change) * Price Effect (price effect)

8.10 The Hicks-Moorsteen TFP Index (HMB Index)

The Hicks-Moorsteen TFP index, proposed byBjurek (1996), is a method for calculating a TFP index and is therefore also called the HMB index (Hicks-Moorsteen-Bjurek).

Compared with the wide application of the Malmquist index, the HMB index has been applied very little. There may be three reasons. First, the HMB index is more complex to compute: the Malmquist index involves four components, whereas the HMB index involves eight, doubling the computational complexity. Second, the HMB index is more difficult to decompose. Third, no dedicated software provides the HMB index and its full decomposition.

One advantage of the HMB index is that it measures the TFP index from both the input and the output side, so no model orientation has to be set.

MaxDEA 9 has made two contributions to promoting the application of the HMB index: first, it extends the HMB index from adjacent benchmarking to all benchmarking methods supported by the Malmquist index, such as fixed benchmarking, global benchmarking, sequential benchmarking and window benchmarking; second, it provides the same decomposition method as for the Malmquist index. From the user's point of view, the HMB index can now be used exactly as the Malmquist index is used.

The rationale and calculation of the HMB index are not described in detail here; see the help documentation within the software.

8.11 Some Suggestions for Panel Data Analysis

1) Use terms with a clear meaning, such as "technical efficiency", "pure technical efficiency", "scale efficiency", "cost efficiency", "allocative efficiency", "total factor productivity" and "total factor productivity index"; avoid terms with an unclear meaning, such as "efficiency", "static efficiency" and "dynamic efficiency".

2) Avoid terms of the form "factor + efficiency" that are unclear or incorrect, such as "energy efficiency", "energy total factor productivity", "land efficiency" and "land total factor productivity"[i]. Terms with a clear meaning should be used instead, such as "total factor productivity of industry XX", "total factor productivity index of industry XX", "agricultural total factor productivity" and "agricultural total factor productivity index".

3) The technical efficiency of a single period, Score_t(x_t,y_t), is normally used only to calculate EC(t-1,t) when MI(t-1,t) is decomposed. Analysing Score_t(x_t,y_t) on its own, apart from MI(t-1,t), easily leads to misleading conclusions.

4) MI(t-1,t) reflects only the change of one DMU between two adjacent periods. To display the change of a DMU over several consecutive periods, the cumulative MI under global benchmarking, MI(0,t), is recommended; to display the changes of several DMUs over several consecutive periods, Score_g(x_t,y_t) under global benchmarking is recommended.

8.12 The Window DEA Model for Panel Data

Window DEA[ii] treats the data of each DMU in each period as a relatively independent "DMU" (the quotation marks here distinguish it from an actual DMU); if there are n DMUs and p periods, the total number of "DMUs" is np.

If the width of the window is d ($d \le p$), the number of windows is $w = p - d + 1$. The number of "DMUs" in each window is nd, which is d times the number of DMUs in a single period. Increasing the number of "DMUs" is the most obvious function of the window model.

If the window width is set to 1, the number of "DMUs" in a window is the smallest and equals the number of actual DMUs; this amounts to separating the data of the various periods and then analysing the "DMUs" of each period separately. The resulting efficiency score equals Score_t(x_t,y_t) in the adjacent benchmarking (average of two indices) MI model introduced above.

If the window width is set to p, the number of periods contained in the panel data, the number of "DMUs" in a window is the largest; this amounts to pooling the "DMUs" of all periods into one overall reference set and then analysing all "DMUs" against this reference set. The resulting efficiency score equals Score_g(x_t,y_t) in the global benchmarking MI model introduced above.

8.13 How to Analyse the Determinants of TFP When TFP Cannot Be

[i]Energy and land are both factors of production; if the productivity of energy and of land is analysed separately, the appropriate terms are "energy single-factor productivity" and "land single-factor productivity".

[ii]Here Window is mostly rendered in the Chinese DEA literature as shichuang (viewing window), a borrowing from the "Windows operating system". Since the Windows operating system works through visual windows, that rendering is quite apt. In DEA models, however, the window denotes a window period containing several periods and carries no visual connotation, so chuangkou (window) or chuangkou qi (window period) is more appropriate. Models of this type may be called in Chinese "window data envelopment analysis (window DEA)", "window-period data envelopment analysis" or "data envelopment analysis window model".

Measured

DEA methods can calculate only the TFP index (TFP change), not the value of TFP itself. To analyse the determinants of the rate of TFP change, the TFP index can be used directly as the dependent variable of a regression model; to analyse the determinants of TFP itself, the TFP index can in fact also be used as the dependent variable, but note that a particular TFP index (or a particular efficiency score) must be used as the dependent variable and that a panel data fixed-effects regression model must be adopted. The following describes how to use the "TFP index" as the dependent variable to analyse the determinants of "TFP".

Method 1: Using the "logarithm of the cumulative global Malmquist index" as the dependent variable in a fixed-effects regression model yields the same regression coefficients as using the "logarithm of TFP" as the dependent variable.

Method 2: Using the "logarithm of the global efficiency score" as the dependent variable in a fixed-effects regression model yields the same regression coefficients as using the "logarithm of TFP" as the dependent variable.

Readers may derive and prove these results themselves.

The specific calculation procedure is as follows:

Method 1: Use the "logarithm of the cumulative global Malmquist index" in place of the "logarithm of TFP" as the dependent variable Y in the regression.

1) In the panel data model of the MaxDEA software, select "TFP index"; for "benchmarking type" select "c) Global benchmarking"; for "method for calculating and decomposing the TFP index" select "Malmquist index"; for "TFP index for which two periods" select "(t0, t)".

2) Take the logarithm of the "cumulative global Malmquist index", that is, MI(t0, t) in the output of MaxDEA 9.0, as the dependent variable Y.

3) Run the regression with a fixed-effects model; the Stata command is as follows:

Set the panel data: xtset id variable period variable

Fixed-effects model (model options can be added): xtreg dependent variable independent variables, fe

Method 2: Use the "logarithm of the global efficiency score" in place of the "logarithm of TFP" as the dependent variable Y in the regression.

1) In the panel data model of the MaxDEA software, select "TFP index"; for "benchmarking type" select "c) Global benchmarking"; for "method for calculating and decomposing the TFP index" select "Malmquist index" or "Luenberger index", either of which is acceptable, because the regression model of this method does not use the TFP index.

2) Take the logarithm of the global efficiency scores of all periods (including period 0), that is, of Score in the table "Score_g(x_t,y_t)", as the dependent variable Y.

3) Run the regression with a fixed-effects model.

To verify that the regression coefficients of Method 2 are the same as those of Method 1, the data should not include Score_g for period 0 (the first period); otherwise the sample sizes of Method 2 and Method 1 will differ and the results will differ. In practical applications, however, Method 2 may include Score_g for period 0. The dependent variable of Method 1 uses the TFP index and can cover only periods 1

to t, whereas the dependent variable of Method 2 uses the global efficiency score and covers periods 0 to t; Method 2 can therefore draw on one more period of data than Method 1, so using the "logarithm of the global efficiency score" in place of the "logarithm of TFP" as the dependent variable is the better approach.

The method described above uses "global benchmarking" as the benchmarking type; "fixed benchmarking" or "window benchmarking (fixed)" can also be selected. In the output of the latter two benchmarking types, the counterpart of the table "Score_g(x_t,y_t)" is the table "Score_f(x_t,y_t)".

Three regression models can be constructed using the three approaches "global benchmarking", "fixed benchmarking" and "window benchmarking (fixed)" respectively; the goodness of fit of the three models can then be compared, and finally the model with the best fit can be selected.

Chapter 9 Innovative Two-Stage and Generalized Network DEA

As the core non-parametric tool for evaluating the relative efficiency of decision-making units (DMUs), traditional data envelopment analysis (DEA) treats each DMU as a “black box” with an invisible internal structure, focusing only on the transformation from final inputs to final outputs. Real production systems, however, often contain complex sub-processes, multi-stage structures, and intermediates flowing among them. Network DEA was proposed precisely to open this “black box” and to analyze the operating efficiency of each node (stage) within the system and its impact on the system as a whole. In the technical logic of network DEA, the individual production stages are explicitly independent of one another technologically; that is, each node constructs its own production possibility set (PPS). This technical independence makes it feasible and legitimate to form linear combinations of the Stage 1 technology of one DMU with the Stage 2 technology of another.[i]

Although two-stage and generalized network DEA (covering multiple nodes, with serial and parallel structures, and arbitrary topologies) greatly extends the applicability of conventional DEA, it has long faced a pain point at the theoretical and algorithmic level: **the non-uniqueness of the system efficiency decomposition, and the fact that the node efficiency values obtained from the decomposition are often not the efficiency values the nodes would attain under their own stand-alone PPS**. In solving the multiplier model of conventional network DEA, multiple optimal solutions may exist, so the algebraic mapping between the overall system efficiency and the efficiencies of the sub-nodes is unstable. Such an unstable decomposition obstructs further analysis of the node efficiency values of network DEA models (e.g., TFP index analysis at the node level). To address this problem, previous studies have mostly relied on artificially imposing secondary objectives, introducing subjective preference weights, or constructing leader–follower game models to forcibly “lock in” the decomposition results.

The innovative two-stage network DEA model developed by the MaxDEA team fundamentally resolves this difficulty: the system efficiency and the two stage efficiencies naturally satisfy a unique equality relationship, namely **the overall system efficiency equals the product of the two unique stage efficiencies.** On this basis, it is extended to “generalized network DEA” and “generalized network DEA based on full-dimensional efficient facets (FDEF)”.

If you cite this work in academic research or papers, please refer to the following citation formats:

English format: MaxDEA Team. (2026-10-26). *Innovative Two-Stage and Generalized Network DEA*. MaxDEA. https://maxdea.com/2026/202609.pdf.

Chinese format: MaxDEA 团队. 新型两阶段网络 DEA 和广义网络 DEA [EB/OL]. (2026-10-26) [YYYY-MM-DD]. https://maxdea.com/2026/202609.pdf. [YYYY-MM-DD] is the citation (access) date.

[i] In this chapter, "Stage" is used in the two-stage network DEA, and "Node" is used in the generalized network DEA, to represent the internal structure of the network. In addition, for the weight vectors v and u, "weight" and "(shadow) price" are used interchangeably.

Notation and Data Description

Formulas are written in matrix and vector notation. For the convenience of readers who wish to copy the formulas, **plain-text format** is used in most cases.

Matrices: X, Y, and Z are the input, output, and intermediate matrices. In multiplier models the number of rows of the matrices equals the number of sample DMUs; in envelopment models the number of columns equals the number of sample DMUs.

X, Y, Z > 0.

Column vectors: x_o, y_o, z_o denote the input, output, and intermediate vectors of the DMU under evaluation (the subscript $_o$ is used for better visual prominence). λ is the vector of linear-combination coefficients of the DMUs in the envelopment model.

Row vectors: v, u, and w are the weights of the inputs, outputs, and intermediates in multiplier models.

v, u, w, and λ are non-negative variables (≥ 0) by default.

Bold **1** denotes a row or column vector whose elements are all 1.

$u_o \in R$ is a scalar variable representing the returns-to-scale constant term (the intercept) in the VRS multiplier model.

Superscripts: $^{(1)}$ $^{(2)}$ $^{(t)}$ denote Stage 1, Stage 2, and node t; $^{(in)}$ $^{(out)}$ indicate whether the intermediates act as inputs or outputs at node t.

9.1 Two-Stage Network DEA (Constant Returns to Scale, CRS)

9.1.1 The Two-Stage Network DEA Model

Kao & Hwang (2008); Liang, Cook, & Zhu (2008)

Efficiency of Stage 1

$$E^{(1)}* = \max (w \cdot z_o / v \cdot x_o)$$

$$\text{s.t.} \quad (w \cdot Z) / (v \cdot X) \le 1 \qquad (9.22)$$

Efficiency of Stage 2

$$E^{(2)}* = \max (u \cdot y_o / w \cdot z_o)$$

$$\text{s.t.} \ (u \cdot Y) / (w \cdot Z) \le 1 \qquad (9.23)$$

Overall system efficiency

$$E* = E^{(1)} \times E^{(2)} = \max (u \cdot y_o / w \cdot z_o) \times (w \cdot z_o / v \cdot x_o) = \max (u \cdot y_o / v \cdot x_o)$$

$$\text{s.t.} \quad (w \cdot Z) / (v \cdot X) \le 1$$

$$(u \cdot Y) / (w \cdot Z) \le 1 \qquad (9.24)$$

Applying the Charnes-Cooper transformation yields the following linear programming (LP) model (input-oriented):

$$\max u \cdot y_o$$

$$\text{s.t.} \ v \cdot X - w \cdot Z \ge 0$$

$$w \cdot Z - u \cdot Y \ge 0$$

$$v \cdot x_o = 1 \qquad (9.25)$$

Dual envelopment model

$$\min \theta$$

$$\text{s.t.} \ X \cdot \lambda^{(1)} \le x_o \cdot \theta$$

$$Y \cdot \lambda^{(2)} \ge y_o$$

$$Z \cdot \lambda^{(1)} \ge Z \cdot \lambda^{(2)}$$

$$\theta \in R \qquad (9.26)$$

The overall system efficiency of two-stage network DEA can be decomposed into the product of two components:

$$E* = (u \cdot y_o) / (v \cdot x_o) = (u \cdot y_o) / (w \cdot z_o) \cdot (w \cdot z_o) / (v \cdot x_o) = E^{(1)} \times E^{(2)} \le E^{(1)}* \times E^{(2)}*$$

By opening the "black box" of the system, the above two-stage network DEA framework successfully captures the flow of intermediate products within the system. Within this theoretical architecture, the model exhibits the following features and defects:

Feature 1: rigid synergy with equal intermediate prices (Rigid Synergy)

The multiplier model strictly requires that the weights $w^{(out)}$ of the Stage 1 intermediates (as outputs) and the weights $w^{(in)}$ of the Stage 2 intermediates (as inputs) be absolutely equal on a one-to-one basis.

Feature 2: hybrid nature of the technology improvement path

When mapped into the dual envelopment model, the rigid synergy mechanism appears as a mixture of radial and non-radial adjustment: the model imposes radial proportional scaling constraints (θ) only on the initial input x (or the final output y in output-oriented models), while the intermediates z are granted complete freedom. This hybrid structure weakens the economic meaning of the model and causes the projections to deviate from the classical technology improvement path.

Defect 1: non-uniqueness of the system efficiency decomposition

Under this framework, although the overall system efficiency is uniquely determined, the stage efficiency values obtained from the decomposition may not be unique, because the multiplier model often has multiple optimal solutions.

$$E^* = E^{(1)} \times E^{(2)}, \text{ where } E^{(1)} \text{ and } E^{(2)} \text{ may not be unique.}$$

Defect 2: the decomposed stage efficiency values may deviate from their respective PPS

The stage efficiencies obtained under the rigid synergy mechanism may not be the true technical efficiencies of the stages under their own PPS.

$$E^{(1)} \le E^{(1)*},\ E^{(2)} \le E^{(2)*}.$$

9.1.2 Two-Stage Network DEA Model Based on Conservation of the Virtual Value of Intermediates

To address the pain points of the above framework, the following two-stage network DEA model with flexible synergy, which requires only that the total value of the intermediates be equal, is proposed:

$$E^* = \max (u \cdot y_o) / (v \cdot x_o)$$

$$\text{s.t.} \quad (w^{(1)} \cdot Z) / (v \cdot X) \le 1$$

$$(u \cdot Y) / (w^{(2)} \cdot Z) \le 1$$

$$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o} \qquad (9.27)$$

Applying the Charnes-Cooper transformation yields the following linear programming (LP) model (input-oriented):

$E^* = \max u \cdot y_o$

s.t. $v \cdot X - w^{(1)} \cdot Z \geq 0$

$w^{(2)} \cdot Z - u \cdot Y \geq 0$

$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o}$

$v \cdot x_o = 1$ (9.28)

Dual envelopment model

$E^* = \min \theta$

s.t. $X \cdot \lambda^{(1)} \leq x_o \cdot \theta$

$\mathbf{Z \cdot \lambda^{(1)} \geq z_o \cdot \tau}$

$\mathbf{Z \cdot \lambda^{(2)} \leq z_o \cdot \tau}$

$Y \cdot \lambda^{(2)} \geq y_o$

$\theta, \tau \in R$ (9.29)

Alternatively, applying the Charnes-Cooper transformation yields the following linear programming (LP) model (output-oriented):

$1/E^* = \min v \cdot x_o$

s.t. $v \cdot X - w^{(1)} \cdot Z \geq 0$

$w^{(2)} \cdot Z - u \cdot Y \geq 0$

$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o}$

$u \cdot y_o = 1$ (9.30)

Dual envelopment model

$1/E^* = \max \phi$

s.t. $X \cdot \lambda^{(1)} \leq x_o$

$\mathbf{Z \cdot \lambda^{(1)} \geq z_o \cdot \tau}$

$\mathbf{Z \cdot \lambda^{(2)} \leq z_o \cdot \tau}$

$Y \cdot \lambda^{(2)} \geq y_o \cdot \phi$

$\phi, \tau \in R$ (9.31)

The two-stage network DEA model based on conservation of the virtual value of intermediates has the following properties:

1) Flexible synergy of intermediates

It abandons the strong exogenous constraint that the price of each intermediate must be equal, and instead requires only that all intermediates of the DMU under evaluation satisfy “equal total value” as they flow within the system. This matches the circulation characteristic that internal settlement of

intermediates generates no externalities, and allows the intermediate weights (shadow prices) to be determined within the system according to the classical DEA optimization principle.

$$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o$$

2) Purely classical technology improvement path

It fully adopts the classical radial proportional technology improvement path, completely preserving the economic meaning of the classical DEA model.

$$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$$

$$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$$

3) The system efficiency decomposition is unique, and the decomposed stage efficiency values are the technical efficiencies under each stage's own PPS.

9.1.3 Efficiency Decomposition of the Two-Stage Network DEA Model (CRS)

In the two-stage network DEA model, the overall network efficiency obtained from model (9.27) equals the product of the efficiency of stage 1 obtained from model (9.22) and the efficiency of stage 2 obtained from (9.23) **(the proof is given in the appendix)**:

$$E^* = E^{(1)*} \times E^{(2)*}$$

9.1.4 The Flexible Synergy Mechanism for Intermediates

In network DEA models, the shadow prices determined independently by each stage are valid only within that stage, and they can be scaled up or down as a whole without loss of equivalence. Since the production technologies of the two stages are independent of each other, a synergy mechanism is needed when evaluating the overall system efficiency, in order to reflect the bridging role of the intermediates in the network structure and the rules governing their flows.

Flexible synergy abandons the stringent constraint of rigid synergy (one-to-one equality of intermediate prices, $w^{(1)} = w^{(2)}$), and relaxes it to "equality of the total value of the intermediates between the production side and the consumption side within the system" ($w^{(1)} \cdot z_o = w^{(2)} \cdot z_o$). It is precisely this core change that achieves a unique decomposition of system efficiency and keeps the decomposed stage efficiencies consistent with the stand-alone technical efficiencies of the stages.

Analysis from the Perspective of Shadow Prices and Production Function Theory

In duality and linear programming theory, the multipliers can be directly interpreted as the internal shadow prices or marginal valuations of the inputs, outputs, and intermediates.

Rigid synergy forces $w^{(in)} = w^{(out)}$, imposing a unified cross-stage "exogenous market price" on the intermediate products. However, Stage 1 and Stage 2 embody different production technologies, and the

marginal rates of substitution (MRS) of the intermediates differ between the two stages. Forcing the shadow prices of the intermediates to be equal one-to-one across the two stages means fitting two independent convex frontiers with a single linear tangent plane, which inevitably causes the fitted frontier to deviate from the frontiers of the two stages, so that the stage efficiencies are distorted by the rigid synergy constraint.

Flexible synergy only requires value conservation when the intermediates are transferred within the system, which grants $w^{(1)}$ and $w^{(2)}$ independent degrees of freedom so that each can precisely match the normal vector of the optimal frontier of its own stage (i.e., each stage's independent local optimal shadow prices). As a result, the stage technical efficiencies evaluated by the network system are consistent with the technical efficiencies evaluated independently for each stage.

Analysis from the Perspective of Weighted Virtual Input-Output Theory

The core idea of classical DEA is to aggregate multiple inputs and multiple outputs into scalar virtual inputs and virtual outputs, and—under the constraint that the ratio of virtual output to virtual input does not exceed 1—to give the DMU under evaluation complete freedom to seek optimal weights.

Intermediates play a dual role: they are the virtual output of Stage 1 and the virtual input of Stage 2. Rigid synergy requires a single weight vector to match both input-output structures simultaneously, which restricts each stage's freedom in optimization and prevents the model from reaching the stages' optimal solutions. Flexible synergy removes this over-constraint and imposes only the constraint that the virtual aggregated value of the intermediates at their producing side be equal to that at their consuming side. This ensures that the virtual value passed from Stage 1 to Stage 2 is neither created out of nothing nor destroyed without reason—perfectly consistent with the closed-loop characteristic that intermediate products exist only inside the network system and generate no externalities—while retaining the complete freedom of the DMU under evaluation to seek the optimal weights of each stage, so that each stage can still attain its own optimal solution under the network DEA architecture.

Table 9-1 Comparison of the constraints on intermediates between rigid synergy and flexible synergy

	Rigid Synergy	**Flexible Synergy**
Multiplier model	**Rigid** The prices of the intermediates are equal one-to-one across the two stages $w^{(1)} = w^{(2)}$	**Flexible** The total value of the intermediates is equal across the two stages $w^{(1)} \cdot z_o = w^{(2)} \cdot z_o$
Envelopment	**Free scaling**	**Radial scaling**

model	Only the circulation logic must be observed $z_i \cdot \lambda^{(2)} \le \hat{z}_i \le z_i \cdot \lambda^{(1)}, \forall_i$ Each intermediate is scaled freely and independently	Complying with the circulation logic and the radial scaling of the intermediates $z_i \cdot \lambda^{(2)} \le z_{io} \cdot \tau \le z_i \cdot \lambda^{(1)}, \forall_i$ Or, expressed as $z_i \cdot \lambda^{(2)} \le \hat{z}_i \le z_i \cdot \lambda^{(1)}, \forall_i$ $\hat{z}_i = z_{io} \cdot \tau$ τ is the scaling scalar of the intermediates

* When there is only one intermediate in the network, the two synergy mechanisms are equivalent: $w^{(1)} \cdot z_o = w^{(2)} \cdot z_o \rightarrow w^{(1)} = w^{(2)}$.

9.1.5 Why the Efficiency Decomposition of Two-Stage Network Models Takes the Product Form

The multiplicative structure of the two-stage system efficiency decomposition avoids excessive substitution compensation, in which a highly efficient stage would overly compensate for an inefficient one. “Neither stage may fall behind” is an inherent requirement of a two-stage network system. The product form of efficiency embodies the typical property of a serial system: the inefficiency of any sub-stage ($E^{(t)*} < 1$) drags down the overall system efficiency multiplicatively.

For example, two companies A and B produce the same product from the same raw materials. In Stage 1, company A uses 100 units of raw materials to produce 100 units of intermediate products, while company B uses 100 units of raw materials to produce 10 units; compared with A, B’s Stage 1 efficiency is 0.1. In Stage 2, company A uses 100 units of intermediates to produce 100 units of final products, while company B uses 10 units of intermediates to produce 9 units; compared with A, B’s Stage 2 efficiency is 0.9. The product form correctly reflects this serial-chain effect: B’s overall efficiency is 0.09, not the 0.5 given by the arithmetic average form.

9.2 Two-Stage Network DEA (Variable Returns to Scale, VRS)

9.2.1 Model Definitions

1) Input-oriented model (the intercept is assigned to the output side):

Stage 1 model:

$$E^{(1)}* = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o)$$
$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0 \qquad (9.32)$$

Stage 2 model:

$$E^{(2)}* = \max (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o)$$

$$\text{s.t. } w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0 \qquad (9.33)$$

Two-stage network model:

$$E* = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o) \cdot (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o) = \max (u \cdot y_o - u_o^{(2)}) / (v \cdot x_o)$$

$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$$
$$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$$
$$(w^{(1)} \cdot z_o - u_o^{(1)}) = w^{(2)} \cdot z_o \qquad (9.34)$$

Applying the Charnes-Cooper transformation converts it into a linear program.

$$E* = \max (u \cdot y_o - u_o^{(2)})$$

$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$$
$$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$$
$$v \cdot x_o = 1$$
$$w^{(1)} \cdot z_o - u_o^{(1)} = w^{(2)} \cdot z_o \qquad (9.35)$$

Dual envelopment model:

$$E* = \min \theta$$

$$\text{s.t. } X \cdot \lambda^{(1)} \leq x_o \cdot \theta$$

$$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$$

$$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$$

$$Y \cdot \lambda^{(2)} \geq y_o$$

$$\mathbf{1} \cdot \lambda^{(1)} = \tau$$

$$\mathbf{1} \cdot \lambda^{(2)} = 1$$

$$\theta, \tau \in R \qquad (9.36)$$

2) Output-oriented model (the intercept is assigned to the input side):

Stage 1 model:

$E^{(1)}* = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o)$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$ (9.37)

Stage 2 model:

$E^{(2)}* = \max (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o)$

$s.t.\ w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$ (9.38)

Two-stage network model:

$E* = \max (w^{(1)} \cdot z_o) / (v \cdot x_o + u_o^{(1)}) \cdot (u \cdot y_o) / (w^{(2)} \cdot z_o + u_o^{(2)}) = \max (u \cdot y_o) / (v \cdot x_o + u_o^{(1)})$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$

$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$

$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o + u_o^{(2)}$ (9.39)

Applying the Charnes-Cooper transformation converts it into a linear program.

$1/E* = \min (v \cdot x_o + u_o^{(1)})$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$

$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$

$u \cdot y_o = 1$

$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o + u_o^{(2)}$ (9.40)

Dual envelopment model:

$1/E* = \max \phi$

$s.t.\ X \cdot \lambda^{(1)} \leq x_o$

$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$

$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$

$Y \cdot \lambda^{(2)} \geq \phi \cdot y_o$

$\mathbf{1} \cdot \lambda^{(1)} = 1$

$\mathbf{1} \cdot \lambda^{(2)} = \tau$

$\phi, \tau \in R$ (9.41)

9.2.2 Efficiency Decomposition

In the two-stage network DEA model, the overall network efficiency obtained from model (9.34) equals the product of the efficiency of stage 1 obtained from the model (9.32) and the efficiency of stage 2 obtained from model (9.33) (The proof is similar to that of the CRS model):

$$\mathbf{E* = E^{(1)}* \times E^{(2)}*}$$

9.3 Generalized Network DEA Model

9.3.1 Model Definition

$E^* = \max \Sigma((u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) / (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)})) / N$

s.t. $v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + (\eta_3^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) = 0$ (9.42)

N is the number of nodes in the network

η indicates how the intercept term is assigned in the VRS model: $\eta_1^{(t)} + \eta_2^{(t)} + \eta_3^{(t)} + \eta_4^{(t)} = 1, \eta \in \{0, 1\}, \forall t$

If node (t) has no input $x^{(t)}$: $\eta_1^{(t)} = 0$

If node (t) has no output $y^{(t)}$: $\eta_2^{(t)} = 0$

If node (t) has no intermediates $z^{(in)(t)}$: $\eta_3^{(t)} = 0$

If node (t) has no intermediates $z^{(out)(t)}$: $\eta_4^{(t)} = 0$

In the input-oriented model, the intercept is assigned to the output side: $\eta_2^{(t)} + \eta_4^{(t)} = 1$, if node (t) has output $y^{(t)}$, $\eta_2^{(t)} = 1$

In the output-oriented model, the intercept is assigned to the input side: $\eta_1^{(t)} + \eta_3^{(t)} = 1$, if node (t) has input $x^{(t)}$, $\eta_1^{(t)} = 1$

CRS model: all η are 0

9.3.2 Solving the Model

First solve the stand-alone node models to obtain each node's optimal efficiency value $E^{(t)*}$:

$E^{(t)*} = \max (u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) / (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)})$

s.t. $v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0$ (9.43)

The objective function of the generalized network model is a sum of fractions, which makes the model a nonlinear program, and it needs to allow the node weights to be scaled proportionally; it cannot be linearized via the Charnes-Cooper transformation. It can be solved by the Dinkelbach parametric approach.

Parametric model (input-oriented; $\theta^{(t)}$ are the model parameters):

$D^* = \max \Sigma((u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) - \theta^{(t)} \cdot (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)}))$

s.t. $v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} \geq 1, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + \eta_4^{(t)} \cdot u_o^{(t)}) = 0$ (9.44)

Parametric model (output-oriented; $\phi^{(t)}$ are the model parameters):

$D^* = \max \Sigma(\phi^{(t)} \cdot (u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) - (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)}))$

s.t. $v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} \geq 1, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + \eta_3^{(t)} \cdot u_o^{(t)}) = 0$ (9.45)

Taking the input-oriented model as an example, set the model parameters $\theta^{(t)}$ for each node equal to its efficiency $E^{(t)*}$ obtained from the stand-alone node model (9.43). Solve the model; if the optimal

objective function value D* = 0, the network exhibits **lossless synergy**; if D* < 0 (D* must be ≤ 0), the network exhibits **lossy synergy**.

Alternatively, the dual model of the parametric model can be solved:

$D^* = \max \Sigma(\alpha^{(t)})$

s.t. $X^{(t)} \cdot \lambda^{(t)} \le x_o^{(t)} \cdot (\theta^{(t)} - \alpha^{(t)}), \forall t$

$Z^{(in)(t)} \cdot \lambda^{(t)} \le z_o^{(in)(t)} \cdot (\theta^{(t)} - \alpha^{(t)} - \gamma), \forall t$

$Y^{(t)} \cdot \lambda^{(t)} \ge y_o^{(t)}, \forall t$

$Z^{(out)(t)} \cdot \lambda^{(t)} \ge z_o^{(out)(t)} \cdot (1 - \gamma), \forall t$

$\gamma \in R, \alpha^{(t)} \ge 0$ (9.46)

VRS model:

If node (t) has output $y^{(t)}$: $\mathbf{1} \cdot \lambda^{(t)} = 1, \forall t$

If node (t) has no output $y^{(t)}$: $\mathbf{1} \cdot \lambda^{(t)} = 1 - \gamma, \forall t$

9.3.3 Method for Computing System Efficiency

Taking the input-oriented model as an example:

1) The model with **lossless synergy (**D* = 0)

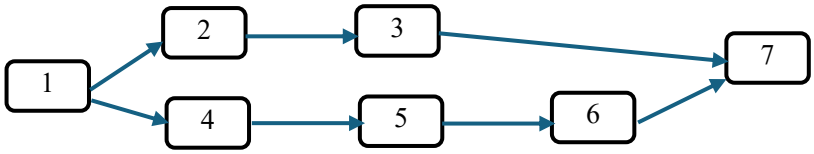
The node efficiency values $E^{(t)*}$ obtained from the stand-alone node models are equal to the node efficiency values computed by the network model:

When the model can achieve **lossless synergy**, the network model's objective function can be any monotonically increasing function of the node efficiencies $f(E^{(1)}, \dots, E^{(N)})$, and the stand-alone optimal solutions of the nodes can all be combined to construct an optimal solution of the network model. Accordingly, we define the following universal method for computing the overall system efficiency (power-product form):

$$E^* = \prod_{t=1}^{N} (E^{(t)*})^{\omega^{(t)}}, \ 1 \le \sum \omega^{(t)} \le N$$

$\omega^{(t)}$ reflects the role of each node in the network: **if $\omega^{(t)} = 1, \forall t$, then the system efficiency is the product of all node efficiencies; if $\omega^{(t)} = 1/N$, then the system efficiency is the geometric mean of all node efficiencies; if the $\omega^{(t)}$ values of the nodes are not identical, but $\Sigma\omega^{(t)} = 1$, then the system efficiency is the weighted geometric mean of all node efficiencies.**

The product form applies to serial structures, where the inefficiency of any sub-stage ($E^{(t)*} < 1$) drags down the overall system efficiency multiplicatively; the geometric mean form applies to parallel structures. For network structures containing complex serial and parallel combinations, the following is a simple example of determining the node weights:



(1) First assign the initial weight 1 to the nodes:

$\omega^{(1)} = 1, \omega^{(2)} = 1, \omega^{(3)} = 1, \omega^{(4)} = 1, \omega^{(5)} = 1, \omega^{(6)} = 1, \omega^{(7)} = 1.$

(2) Branch 1 (nodes 2 and 3) and Branch 2 (nodes 4, 5, 6) are in parallel; take the geometric mean of the two branches (the weights of the nodes inside each branch are halved relative to their original values):

$\omega^{(1)} = 1, \omega^{(2)} = 0.5, \omega^{(3)} = 0.5, \omega^{(4)} = 0.5, \omega^{(5)} = 0.5, \omega^{(6)} = 0.5, \omega^{(7)} = 1.$

$\sum \omega^{(t)} = 4.5$, which lies between the geometric mean ($\sum \omega^{(t)} = 1$) and the product ($\sum \omega^{(t)} = 7$).

The above weights are derived only from the structural features of the network. In empirical studies, the final weights can also be determined by taking into account the importance of each node to the system.

2) The model with **lossy synergy (**$D^* < 0$)

$D^* < 0$ indicates lossy synergy in the network, and the node efficiency values under lossy synergy are not unique:

$E^{(t)-} = (u^{(t)*} \cdot y_o^{(t)} + w^{(out)(t)*} \cdot z_o^{(out)(t)} - u_o^{(t)*}) / (v^{(t)*} \cdot x_o^{(t)} + w^{(in)(t)*} \cdot z_o^{(in)(t)}) \leq E^{(t)*}$,

If the dual model (9.46) is solved, then $E^{(t)-} = E^{(t)*} - \alpha^{(t)*} \leq E^{(t)*}$.

The system efficiency under lossy synergy is

$$E^- = \prod_{t=1}^{N} (E^{(t)-})^{\omega^{(t)}}, \ 1 \leq \sum \omega^{(t)} \leq N$$

The node efficiency values obtained under lossy synergy **may not be unique**.

3) Solutions for lossy synergy

One solution is to jointly apply the weight ratio constraint model (also known as cone-ratio DEA, constrained cone DEA, assurance region DEA) to prevent the weights of one or more intermediates from being 0; with properly specified weight ratio constraints, lossless synergy can be achieved.

Another approach is the generalized network DEA based on full-dimensional efficient facets (FDEF) introduced later. In this model all weights are strictly greater than 0, and lossless synergy can be achieved by merely satisfying one simple condition (sufficient but not necessary): the network contains a node that only produces intermediates without consuming any (with $z^{(out)}$ and without $z^{(in)}$), and there is also a node that does not produce intermediates and only consumes them (with $z^{(in)}$ and without $z^{(out)}$). The two-stage network model satisfies this condition.

9.3.4 Compatibility with the Preceding Two-Stage Model

The two-stage network model is a special case of the generalized network model with a two-stage structure, and it satisfies the sufficient condition for **lossless synergy**; when the product form is used to compute the overall system efficiency, although the optimal solutions of the two models (the weights v, u, w) may differ, the resulting node efficiencies and system efficiency are the same.

9.4 Generalized Network DEA Based on Full-Dimensional Efficient Facets (FDEF)

A widely cited defect of classical DEA models is that efficiency measurement does not account for the input and output slacks that often appear in the models. For this reason, various non-radial models that measure efficiency directly based on slacks have been proposed. The input and output slack issue in the model is not a defect of classical DEA itself; rather, it arises because the data sample itself fails to establish a strongly efficient frontier covering the projection region of all samples, so the radial projection point falls on a weakly efficient frontier parallel to at least one coordinate axis. As shown in the figure below, the projection of DMU A is A_1, which lies on the weakly efficient frontier parallel to the vertical axis, and the slack of input x_2 is A_1B. The projection points of the various non-radial models lie on the strongly efficient frontier segment BC. Although non-radial models solve the problem that classical efficiency measurement ignores slacks, they introduce new problems: first, they weaken the economic meaning of the classical dual multiplier model; second, the projection points are often illogical—for example, the additive and SBM models determine the projection point by finding the "farthest distance", which is the least favorable to the DMU under evaluation. Taking the situation shown in the figure below as an example, in the additive and SBM models the projection point is C, which is farther from A, rather than the nearer B.

The slacks appearing in the classical envelopment model correspond to zero weights in the multiplier model. Another approach to addressing the slack problem in the classical model is the full-dimensional efficient facet (FDEF). In the example figure, segment BC of the frontier is an FDEF, and the radial projection based on the affine hull of BC (a straight line in two-dimensional coordinates, a hyperplane in multidimensional space) is A_2, a more reasonable projection point compared to C.

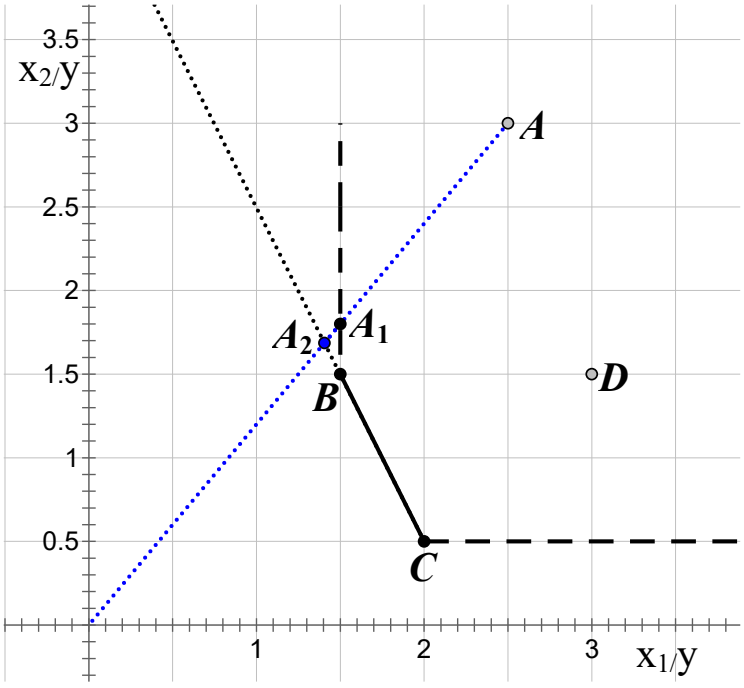


Figure 9-1 Illustration of projections (a CRS model with two inputs and one output)

This section presents a generalized network DEA model based on the affine hull of FDEFs, which eliminates the slack problems present in the classical model, including input and output slacks as well as slacks of the intermediates. The elimination of intermediate slacks guarantees that the projected value of the intermediates as Stage 1 outputs exactly equals the projected value of the intermediates as Stage 2 inputs.

9.4.1 Introduction to Full-Dimensional Efficient Facets (FDEF)

In the convex-envelopment geometric theory of data envelopment analysis (DEA), full-dimensional efficient facets (Full-Dimensional Efficient Facets, FDEF) are among the core concepts characterizing the frontier structure of the production possibility set (PPS).

1. Definition of Full-Dimensional Efficient Facets (FDEF)

In the DEA model containing m inputs and s outputs, the production possibility set $\text{PPS} \subset \mathbb{R}^{m+s}$ is a polyhedron. If an efficient facet formed where an efficient hyperplane on the boundary of this polyhedron intersects it has a geometric dimension exactly equal to $m+s-1$ (i.e., the highest possible degrees of freedom of the space), then the facet is called a **full-dimensional efficient facet (FDEF)**(Olesen & Petersen, 1996, 2003). From the multiplier (dual) perspective, all input multipliers v and output multipliers u corresponding to the normal vector of the hyperplane supporting the FDEF are strictly greater than zero (i.e., $v>0, u>0$).

2. Advantages of the DEA model based on the FDEF affine hull

DEA models constructed with FDEF-based constraints or identification have the following notable advantages:

Complete elimination of the zero-weight problem: conventional DEA models often exhibit the "zero weight" phenomenon due to multiplier degeneracy (i.e., some inputs or outputs are assigned zero weight), whereas FDEF ensures that all indicators play a role in the marginal evaluation.

Sound economic interpretation (marginal rates of substitution): since the normal vector of an FDEF is entirely positive, the marginal rates of substitution (MRS) derived from it are all finite, non-zero, and determinate, avoiding the loss of economic meaning on lower-dimensional efficient facets (LDEF), where MRS values are 0 or infinite.

Improved projection and target setting: projecting an inefficient decision-making unit (DMU) onto an FDEF hyperplane ensures that the improvement targets are strongly efficient, and free of slacks.

3. Methods for identifying all FDEFs in the sample of a DEA model

Methods for solving and identifying all FDEFs on the frontier of a polyhedron currently fall into two main categories:

Computational geometry algorithms based on the convex hull: using vertex enumeration and facet-generation algorithms from computational geometry, a convex hull is constructed over the set of DMU vertices, and the hyperplanes of dimension $m+s-1$ whose normal vectors are greater than 0 are extracted(Olesen & Petersen, 2003).

Iterative search methods based on mixed-integer programming (MIP): using duality theory, MIP is used to search for all hyperplanes with weights strictly greater than zero (Olesen & Petersen, 2003; Jahanshahloo et al., 2007; Jahanshahloo et al., 2009; Olyaie, Roshdi, Jahanshahloo, & Asgharian, 2014).

4. Absence of FDEFs in the sample and severe frontier degeneracy

If a data set yields no FDEF at all when solved, this geometrically means that the envelopment frontier of the sample consists entirely of lower-dimensional efficient facets (LDEF). This carries a clear warning for both theory and application: it indicates severe frontier degeneracy, severely ill-conditioned data, or the curse of dimensionality (i.e., the sample size is severely insufficient relative to the dimensionality of the input-output indicators), causing the production frontier to “fold” or “collapse” into a lower-dimensional subspace in the high-dimensional space(Olesen & Petersen, 1996).

An effective remedy is to introduce a sufficient number of benchmark DMUs as the reference set of the model.

The existence of an FDEF in the sample does not mean that a satisfactory evaluation model can be built. If the model has only one or very few FDEFs, the DMU under evaluation may be projected onto a region far from the FDEF (lying on the supporting hyperplane of the FDEF but far away from the FDEF itself), yielding unrealistic target values. How to judge whether the FDEFs in a sample are sufficient is a question that still needs to be explored in both theory and practice. A practical way to judge the sufficiency of FDEFs is to check whether the projected values (targets) exceed a reasonable range.

9.4.2 Generalized Network DEA Model Based on the Affine Hull of FDEFs

A DEA model based on the affine hull of FDEFs restricts the weights of the model to come from the set of FDEF hyperplanes. The generalized network DEA model with an FDEF-extended frontier is expressed as:

$$E^* = \max \Sigma((u^{(t)}\cdot y_o^{(t)} + w^{(out)(t)}\cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)})\cdot u_o^{(t)}) / (v^{(t)}\cdot x_o^{(t)} + w^{(in)(t)}\cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)})\cdot u_o^{(t)})) / N$$

$$\text{s.t. } \Sigma(w^{(in)(t)}\cdot z_o^{(in)(t)} - w^{(out)(t)}\cdot z_o^{(out)(t)} + (\eta_3^{(t)} + \eta_4^{(t)})\cdot u_o^{(t)}) = 0$$

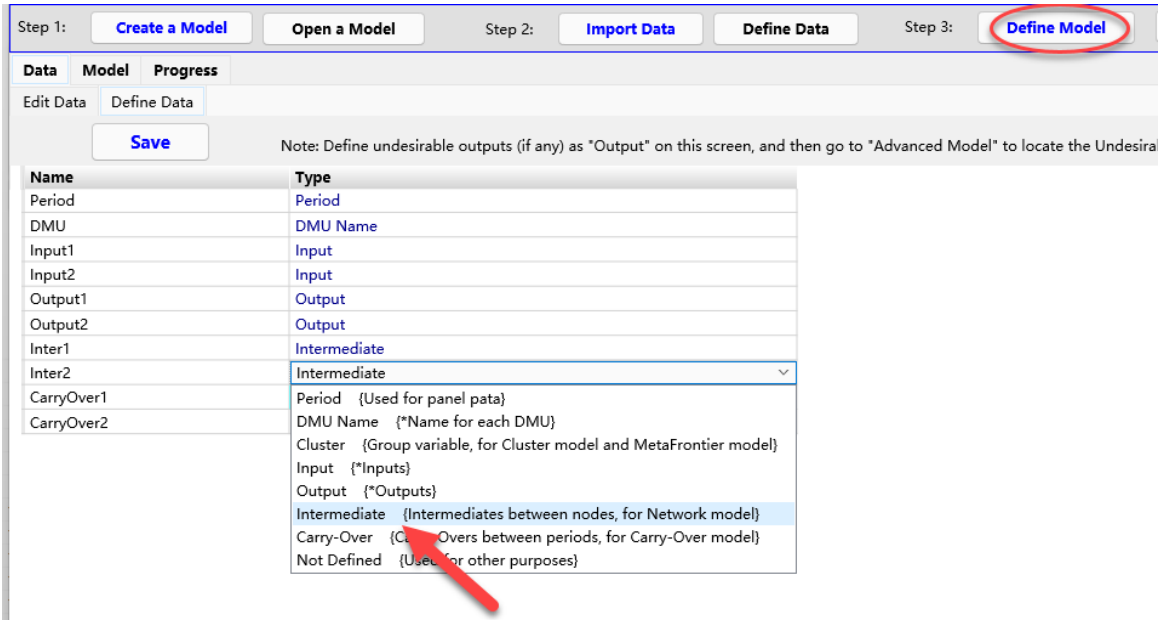
$$(v^{(t)}, u^{(t)}, w^{(in)(t)}, w^{(out)(t)}, u_o^{(t)}) \in F^{(t)} \qquad (9.47)$$

$F^{(t)}$ denotes the set of FDEF hyperplanes of stage t.

In the generalized network DEA model based on the affine hull of FDEFs, since all weights are strictly greater than 0, simply satisfying the following simple condition (sufficient but not necessary) achieves lossless synergy and unique decomposition of system efficiency: the network contains a node that only produces intermediates without consuming any (with $z^{(out)}$ and without $z^{(in)}$), and there is also a node that does not produce intermediates and only consumes them (with $z^{(in)}$ and without $z^{(out)}$).

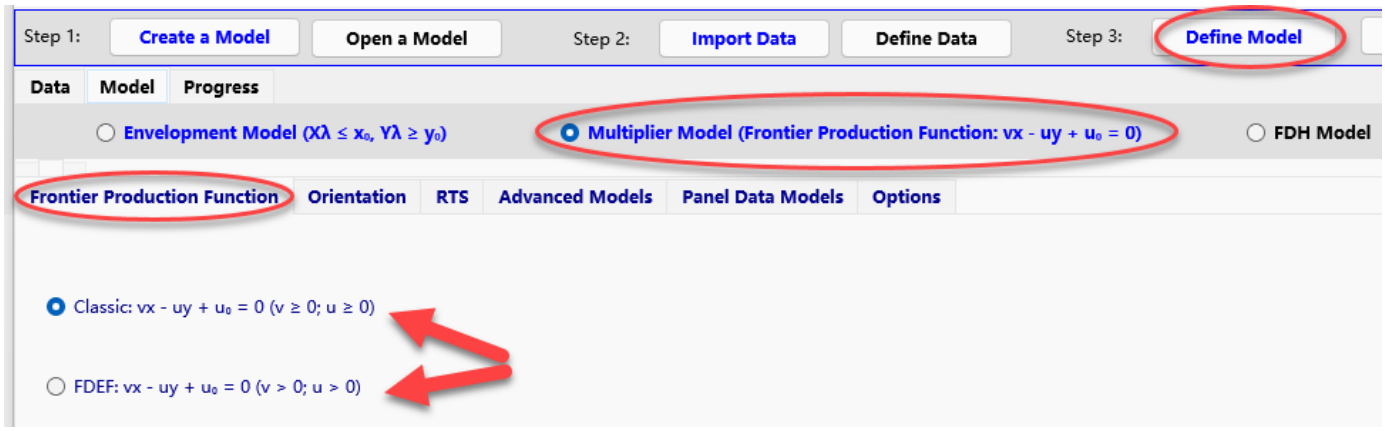
9.5 Software Demonstration of Network DEA

9.5.1 Defining the Data



9.5.2 Setting the General Options of the DEA Model

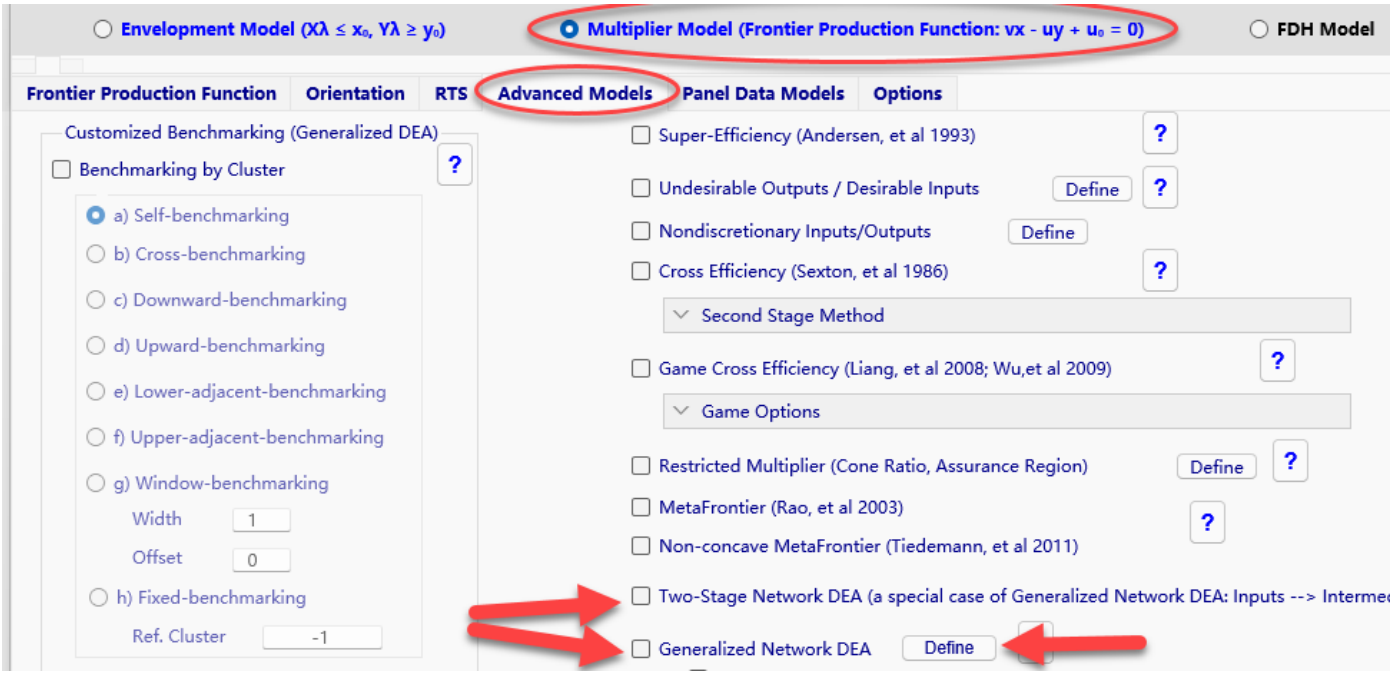
The production function type (classical frontier production function or FDEF affine-hull frontier production function), the model Orientation and RTS.



9.5.3 Setting Other Model Options

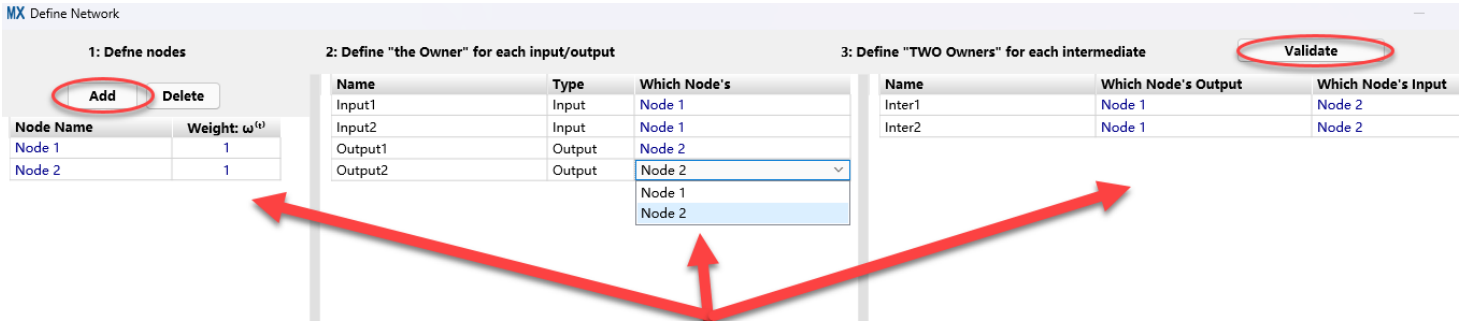
Select **Two-Stage Network DEA** or **Generalized Network DEA**.

Network DEA can be combined with all **panel data models**, including: Malmquist TFP index, HMB TFP index, window DEA, and Carry-Over DEA.



9.5.4 Defining the Network Structure

If "Generalized Network DEA" is selected, the network structure also needs to be defined. For each node of the generalized network model, the initial default weight $\omega^{(t)} = 1$, i.e., the system efficiency is the product of the node efficiencies; if the system efficiency is required to be the geometric mean of the node efficiencies, set the sum of the node weights equal to 1, i.e., $\Sigma\omega^{(t)} = 1$.



9.5.5 Results

The results include the overall system efficiency values, the weights of the inputs, outputs, and intermediates, the projected values (targets) of the inputs, outputs, and intermediates, the node efficiency values, the benchmarks of the nodes and their coefficients (Lambda), etc.; the VRS models also include the returns-to-scale constant term ($u_0^{(t)}$), and the generalized network model also includes D* (used to determine whether the node efficiency values given by the network model are unique).

Appendix: Proof of the Multiplicative Decomposition of Efficiency in the CRS Two-Stage Network DEA

In the two-stage network DEA model, the overall network efficiency equals the product of the optimal stage efficiencies obtained from the Stage 1 and Stage 2 models:

$$E^* = E^{(1)*} \times E^{(2)*}$$

1. **Proof of** $e^* \le e^{(1)*} \times e^{(2)*}$

Let $(v', w^{(1)r}, w^{(2)r}, u')$ be an arbitrary **feasible solution of the overall system model**.

Using the intermediate constraint $w^{(1)r} z_0 = w^{(2)r} z_0$ of the network model, the objective function value corresponding to this feasible solution can be decomposed as:

$$\frac{u'y_0}{v'x_0} = \frac{w^{(1)r} z_0}{v'x_0} \cdot \frac{u'y_0}{w^{(2)r} z_0}$$

By the definition of feasible solutions:

1. The vector pair $(v', w^{(1)r})$ satisfies $\frac{w^{(1)r} z_j}{v'x_j} \le 1\ (\forall j)$, and therefore $(v', w^{(1)r})$ is a feasible solution of the Stage 1 model. By the definition of the maximum of $e^{(1)*}$:

$$\frac{w^{(1)r} z_0}{v'x_0} \le e^{(1)*}$$

2. The vector pair $(w^{(2)r}, u')$ satisfies $\frac{u'y_j}{w^{(2)r} z_j} \le 1\ (\forall j)$, and therefore $(w^{(2)r}, u')$ is a feasible solution of the Stage 2 model. By the definition of the maximum of $e^{(2)*}$:

$$\frac{u'y_0}{w^{(2)r} z_0} \le e^{(2)*}$$

Substituting the above two inequalities into the product decomposition yields:

$$\frac{u'y_0}{v'x_0} \le e^{(1)*} \times e^{(2)*}$$

Since the above inequality holds for **any** feasible solution in the feasible region of the network model, it also holds for the maximum over the feasible region (i.e., the optimal solution):

$$e^* \le e^{(1)*} \times e^{(2)*}$$

2. **Proof of** $e^* \ge e^{(1)*} \times e^{(2)*}$

Let $(v^*, w^{(1)*})$ be an **optimal solution of the Stage 1 model**, satisfying:

$$e^{(1)*} = \frac{w^{(1)*} z_0}{v^* x_0}$$

Let $(w^{(2)*}, u^*)$ be an **optimal solution of the Stage 2 model**, satisfying:

$$e^{(2)*} = \frac{u^* y_0}{w^{(2)*} z_0}$$

Construct a positive scalar scaling factor:

$$\alpha = \frac{w^{(1)*} z_0}{w^{(2)*} z_0} > 0$$

Using the homogeneity of the Stage 2 fractional program (i.e., scaling the numerator and denominator weights proportionally does not change the constraints or the ratio), the Stage 2 optimal solution is linearly scaled to construct new weight vectors:

$$\tilde{w}^{(2)} = \alpha w^{(2)*}, \quad \tilde{u} = \alpha u^*$$

Verify the feasibility of the combined solution $(v^*, w^{(1)*}, \tilde{w}^{(2)}, \tilde{u})$ in the network model:

1. **The Stage 1 constraint**: $\frac{w^{(1)*} z_j}{v^* x_j} \le 1\ (\forall j)$ is clearly satisfied.
2. **The Stage 2 constraint**: for any j, we have:

$$\frac{\tilde{u}y_j}{\tilde{w}^{(2)}z_j} = \frac{\alpha u^* y_j}{\alpha w^{(2)*} z_j} = \frac{u^* y_j}{w^{(2)*} z_j} \leq 1$$

The constraints are satisfied.

3. **The intermediate flow equality constraint**:

$$\tilde{w}^{(2)}z_0 = (\alpha w^{(2)*})z_0 = \left(\frac{w^{(1)*}z_0}{w^{(2)*}z_0}\right)(w^{(2)*}z_0) = w^{(1)*}z_0$$

The coupling constraint $w^{(1)}z_0 = w^{(2)}z_0$ is perfectly satisfied.

Therefore, $(v^*, w^{(1)*}, \tilde{w}^{(2)}, \tilde{u})$ is a valid **feasible solution of the network model**.

Compute the objective function value of this feasible solution in the network model:

$$\frac{\tilde{u}y_0}{v^* x_0} = \frac{\alpha u^* y_0}{v^* x_0} = \frac{u^* y_0}{w^{(2)*} z_0} \cdot \frac{w^{(1)*} z_0}{v^* x_0} = e^{(2)*} \times e^{(1)*}$$

Since e^* is the global maximum of the network model, and the above combined solution is merely one particular solution in its feasible region, it must hold that:

$$e^* \geq e^{(1)*} \times e^{(2)*}$$

3. Conclusion

Combining the inequalities of the upper and lower bounds:

$e^* \leq e^{(1)*} \times e^{(2)*}$ and $e^* \geq e^{(1)*} \times e^{(2)*}$

By the antisymmetry of the order relation on real numbers, it necessarily follows that:

$$e^* = e^{(1)*} \times e^{(2)*}$$

References

Ali, A. I., & Seiford, L. M. (1993). Computational Accuracy and Infinitesimals in Data Envelopment Analysis. *Infor, 31*(4), 290–297.

Amirteimoori, A., & Kordrostami, S. (2012). Generating strong defining hyperplanes of the production possibility set in data envelopment analysis. [Article]. *Applied Mathematics Letters, 25*(3), 605–609. doi: 10.1016/j.aml.2011.09.068

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. [Article]. *Management Science, 39*(10), 1261–1265.

Ando, K., Kai, A., Maeda, Y., & Sekitani, K. (2012). Least distance based inefficiency measures on the Pareto-efficient frontier in DEA. *Journal of the Operations Research Society of Japan, 55*(1), 73–91.

Ando, K., Minamide, M., Sekitani, K., & Shi, J. (2017). Monotonicity of minimum distance inefficiency measures for Data Envelopment Analysis. *European Journal of Operational Research, 260*(1), 232–243. doi: https://doi.org/10.1016/j.ejor.2016.12.028

Aparicio, J., Cordero, J. M., & Pastor, J. T. (2017). The determination of the least distance to the strongly efficient frontier in Data Envelopment Analysis oriented models: Modelling and computational aspects. *Omega, 71*, 1–10. doi: https://doi.org/10.1016/j.omega.2016.09.008

Aparicio, J., & Pastor, J. T. (2013). A well-defined efficiency measure for dealing with closest targets in DEA. *Applied Mathematics and Computation, 219*(17), 9142–9154. doi: https://doi.org/10.1016/j.amc.2013.03.042

Aparicio, J., & Pastor, J. T. (2014). Closest targets and strong monotonicity on the strongly efficient frontier in DEA. *Omega, 44*, 51–57. doi: https://doi.org/10.1016/j.omega.2013.10.001

Aparicio, J., Ruiz, J. L., & Sirvent, I. (2007). Closest targets and minimum distance to the Pareto-efficient frontier in DEA. *Journal of Productivity Analysis, 28*(3), 209–218. doi: DOI 10.1007/s11123-007-0039-5

Asmild, M., Paradi, J. C., Aggarwall, V., & Schaffnit, C. (2004). Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. *Journal of Productivity Analysis, 21*(1), 67–89.

Banker, R. D., & Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. *European Journal of Operational Research, 175*(2), 1311–1320. doi: DOI: 10.1016/j.ejor.2005.06.028

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. [Article]. *Management Science, 30*(9), 1078–1092.

Banker, R. D., & Morey, R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. [Article]. *Operations Research, 34*(4), 513–521.

Banker, R. D., & Thrall, R. M. (1992). Estimation of Returns to Scale Using Data Envelopment Analysis. *European Journal of Operational Research, 62*(1), 74–84.

Barros, C. P., Managi, S., & Matousek, R. (2012). The technical efficiency of the Japanese banks: Non-radial directional performance measurement with undesirable output. *Omega, 40*(1), 1–8. doi: https://doi.org/10.1016/j.omega.2011.02.005

Berg, S. A., Forsund, F. R., & Jansen, E. S. (1992). Malmquist Indexes of Productivity Growth during the Deregulation of Norwegian Banking, 1980-89. *Scandinavian Journal of Economics, 94*, S211–S228.

Bjurek, H. (1996). The Malmquist total factor productivity index. [Article]. *Scandinavian Journal of Economics, 98*(2), 303–313.

Boljuncic, V. (2006). Sensitivity analysis of an efficient DMU in DEA model with variable returns to scale (VRS). *Journal of Productivity Analysis, 25*(1), 173–192. doi: DOI 10.1007/s11123-006-7139-5

Briec, W. (1999). Holder distance function and measurement of technical efficiency. *Journal of Productivity Analysis, 11*(2), 111–131.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic-theory of index numbers and the measurement of input, output, and productivity. *Econometrica, 50*(6), 1393–1414.

Chambers, R. G., Färe, R., & Grosskopf, S. (1996). Productivity growth in APEC countries. *Pacific Economic Review, 1*(3), 181–190. doi: 10.1111/j.1468-0106.1996.tb00184.x

Charnes, A., Clark, C. T., Cooper, W. W., & Golany, B. (1985). A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces. *Annals of Operations Research, 2*(1), 95–112. doi: 10.1007/bf01874734

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Charnes, A., Rousseau, J. J., & Semple, J. H. (1996). Sensitivity and stability of efficiency classifications in data envelopment analysis. *Journal of Productivity Analysis, 7*(1), 5–18.

Chen, C. M. (2013). A critique of non-parametric efficiency analysis in energy economics studies. *Energy*

Economics, 38, 146–152. doi: DOI 10.1016/j.eneco.2013.03.009
Cheng, G., Zervopoulos, P., & Qian, Z. (2013). A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research, 225*(1), 100–105. doi: 10.1016/j.ejor.2012.09.031
Cheng, G., & Zervopoulos, P. D. (2014). Estimating the technical efficiency of health care systems: A cross-country comparison using the directional distance function. *European Journal of Operational Research, 238*(3), 899–910. doi: http://dx.doi.org/10.1016/j.ejor.2014.05.007
Chung, Y. H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management, 51*(3), 229–240.
Cook, W. D., Seiford, L. M., & Zhu, J. (2004). Models for performance benchmarking: measuring the effect of e-business activities on banking performance. *Omega, 32*(4), 313–322. doi: http://dx.doi.org/10.1016/j.omega.2004.01.001
Cooper, W. W., Li, S., Seiford, L. M., Tone, K., Thrall, R. M., & Zhu, J. (2001). Sensitivity and stability analysis in DEA: Some recent developments. *Journal of Productivity Analysis, 15*(3), 217–246.
Cooper, W. W., Park, K. S., & Pastor, J. T. (1999). RAM: A range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. *Journal of Productivity Analysis, 11*(1), 5–42.
Cooper, W. W., Pastor, J. T., Borras, F., Aparicio, J., & Pastor, D. (2011). BAM: a bounded adjusted measure of efficiency for use with bounded additive models. [journal article]. *Journal of Productivity Analysis, 35*(2), 85–94. doi: 10.1007/s11123-010-0190-2
Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-Solver software* (2nd ed.). New York: Springer Science + Business Media.
Emrouznejad, A., & Amin, G. R. (2009). DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling, 33*(1), 486–498. doi: DOI 10.1016/j.apm.2007.11.018
Emrouznejad, A., Anouze, A. L., & Thanassoulis, E. (2010). A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA. *European Journal of Operational Research, 200*(1), 297–304. doi: DOI 10.1016/j.ejor.2009.01.001
Färe, R., Grifell-Tatjé, E., Grosskopf, S., & Knox Lovell, C. A. (1997). Biased Technical Change and the Malmquist Productivity Index. *Scandinavian Journal of Economics, 99*(1), 119–127. doi: 10.1111/1467-9442.00051
Färe, R., & Grosskopf, S. (2003). Nonparametric productivity analysis with undesirable outputs: Comment. *American Journal of Agricultural Economics, 85*(4), 1070–1074.
Färe, R., & Grosskopf, S. (2010). Directional distance functions and slacks-based measures of efficiency. *European Journal of Operational Research, 200*(1), 320–322. doi: DOI 10.1016/j.ejor.2009.01.031
Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis, 3*(1-2), 85–101.
Färe, R., Grosskopf, S., & Lovell, C. A. K. (1983). The structure of technical efficiency. *The Scandinavian journal of economics, 85*(2), 181–190.
Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*: Kluwer-Nijhoff Pub.
Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries. *American Economic Review, 84*(1), 66–83.
Färe, R., & Knox Lovell, C. A. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory, 19*(1), 150–162. doi: Doi: 10.1016/0022-0531(78)90060-1
Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General), 120*(3), 253–290
Fukuyama, H., Maeda, Y., Sekitani, K., & Shi, J. (2014). Input–output substitutability and strongly monotonic p-norm least distance DEA measures. *European Journal of Operational Research, 237*(3), 997–1007. doi: https://doi.org/10.1016/j.ejor.2014.02.033
Fukuyama, H., & Weber, W. L. (2009). A directional slacks-based measure of technical inefficiency. *Socio-Economic Planning Sciences, 43*(4), 274–287. doi: 10.1016/j.seps.2008.12.001
Grosskopf, S. (1986). The Role of the Reference Technology in Measuring Productive Efficiency. *Economic Journal, 96*, 499–513.
Hollingsworth, B., & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters, 10*(11), 733–735. doi: Doi 10.1080/1350485032000133381
Jahanshahloo, G. R., Hosseinzadeh, F., Shoja, N., Sanei, M., & Tohidi, G. (2005). Sensitivity and stability analysis in DEA. *Applied Mathematics and Computation, 169*(2), 897–904. doi: DOI: 10.1016/j.amc.2004.09.092
Jahanshahloo, G. R., Lotfi, F. H., Rezai, H. Z., & Balf, F. R. (2007). Finding strong defining hyperplanes of

production possibility set. *European Journal of Operational Research, 177*(1), 42–54. doi: DOI 10.1016/j.ejor.2005.11.031

Jahanshahloo, G. R., Shirzadi, A., & Mirdehghan, S. M. (2009). Finding strong defining hyperplanes of PPS using multiplier form. *European Journal of Operational Research, 194*(3), 933–938. doi: DOI 10.1016/j.ejor.2008.01.053

Jahanshahloo, G. R., Vakili, J., & Zarepisheh, M. (2012). A Linear Bilevel Programming Problem for Obtaining the Closest Targets and Minimum Distance of a Unit from the Strong Efficient Frontier. *Asia-Pacific Journal of Operational Research, 29*(2). doi: Artn 1250011

Doi 10.1142/S021759591250011x

Kao, C., & Hwang, S.-N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research, 185*(1), 418–429. doi: http://dx.doi.org/10.1016/j.ejor.2006.11.041

Lee, H., & Choi, Y. (2018). Greenhouse gas performance of Korean local governments based on non-radial DDF. *Technological Forecasting and Social Change, 135*, 13–21. doi: https://doi.org/10.1016/j.techfore.2018.07.011

Liang, L., Cook, W. D., & Zhu, J. (2008). DEA models for two-stage processes: Game approach and efficiency decomposition. *Naval Research Logistics, 55*(7), 643–653. doi: Doi 10.1002/Nav.20308

Lotfi, F. H., Jahanshahloo, G. R., & Esmaeili, M. (2007). Sensitivity analysis of efficient units in the presence of non-discretionary inputs. *Applied Mathematics and Computation, 190*(2), 1185–1197. doi: DOI: 10.1016/j.amc.2007.02.002

Lotfi, F. H., Jahanshahloo, G. R., Mozaffari, M. R., & Gerami, J. (2011). Finding DEA-efficient hyperplanes using MOLP efficient faces. *Journal of Computational and Applied Mathematics, 235*(5), 1227–1231. doi: DOI 10.1016/j.cam.2010.08.007

Lovell, C. A. K., & Pastor, J. T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters, 18*(3), 147–151.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de Estadística y de Investigación Operativa, 4*(2), 209–242. doi: 10.1007/bf03006863

Maniadakis, N., & Thanassoulis, E. (2004). A cost Malmquist productivity index. [Article]. *European Journal of Operational Research, 154*(2), 396–409. doi: 10.1016/s0377-2217(03)00177-2

Mirsalehy, A., Abu Bakar, M. R., Lee, L. S., Jaafar, A. B., & Heydar, M. (2014). Directional Slack-Based Measure for the Inverse Data Envelopment Analysis. *The Scientific World Journal, 2014*, 9. doi: 10.1155/2014/138923

Olesen, O. B., & Petersen, N. C. (1996). Indicators of Ill-Conditioned Data Sets and Model Misspecification in Data Envelopment Analysis: An Extended Facet Approach. [research-article]. *Management Science, 42*(2), 205–219.

Olesen, O. B., & Petersen, N. C. (2003). Identification and use of efficient faces and facets in DEA. *Journal of Productivity Analysis, 20*(3), 323–360.

Olyaie, M. D., Roshdi, I., Jahanshahloo, G., & Asgharian, M. (2014). Characterizing and finding full dimensional efficient facets in DEA: a variable returns to scale specification. [research-article]. *The Journal of the Operational Research Society, 65*(9), 1453–1464.

Pastor, J. T., Asmild, M., & Lovell, C. A. K. (2011). The biennial Malmquist productivity change index. *Socio-Economic Planning Sciences, 45*(1), 10–15. doi: http://dx.doi.org/10.1016/j.seps.2010.09.001

Pastor, J. T., & Lovell, C. A. K. (2005). A global Malmquist productivity index. *Economics Letters, 88*(2), 266–271. doi: DOI: 10.1016/j.econlet.2005.02.013

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (1999). An enhanced DEA Russell graph efficiency measure. *European Journal of Operational Research, 115*(3), 596–607.

Portela, M. C. A. S., Thanassoulis, E., & Simpson, G. (2004). Negative data in DEA: a directional distance approach applied to bank branches. *Journal of the Operational Research Society, 55*(10), 1111–1121. doi: DOI 10.1057/palgrave.jors.2601768

Ray, S. C. (2008). The directional distance function and measurement of super-efficiency: an application to airlines data. *Journal of the Operational Research Society, 59*(6), 788–797. doi: DOI 10.1057/palgrave.jors.2602392

Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *American Economic Review, 87*(5), 1033–1039.

Sharp, J. A., Meng, W., & Liu, W. (2007). A modified slacks-based measure model for data envelopment analysis with 'natural' negative outputs and inputs. *Journal of the Operational Research Society, 58*(12), 1672–1677.

Shestalova, V. (2003). Sequential Malmquist indices of productivity growth: An application to OECD industrial activities. *Journal of Productivity Analysis, 19*(2-3), 211–226.

Thore, S., Kozmetsky, G., & Phillips, F. (1994). Dea of Financial-Statements Data - the United-States Computer

Industry. *Journal of Productivity Analysis, 5*(3), 229–248.
Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research, 130*(3), 498–509. doi: Doi: 10.1016/s0377-2217(99)00407-5
Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. [Article]. *European Journal of Operational Research, 143*(1), 32–41.
Tone, K. (2010). Variations on the theme of slacks-based measure of efficiency in DEA. *European Journal of Operational Research, 200*(3), 901–907. doi: DOI 10.1016/j.ejor.2009.01.027
Tone, K., & Tsutsui, M. (2010). An epsilon-based measure of efficiency in DEA - A third pole of technical efficiency. *European Journal of Operational Research, 207*(3), 1554–1563. doi: DOI 10.1016/j.ejor.2010.07.014
Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts, and urban transit. *Journal of Productivity Analysis, 4*(1-2), 183–210. doi: 10.1007/BF01073473
Wilson, P. W. (1995). Detecting Influential Observations in Data Envelopment Analysis. *Journal of Productivity Analysis, 6*(1), 27–45.
Yao, X., Zhou, H., Zhang, A., & Li, A. (2015). Regional energy efficiency, carbon emission performance and technology gaps in China: A meta-frontier non-radial directional distance function analysis. *Energy Policy, 84*, 142–154. doi: https://doi.org/10.1016/j.enpol.2015.05.001
Zhang, N., Zhou, P., & Choi, Y. (2013). Energy efficiency, CO2 emission performance and technology gaps in fossil fuel electricity generation in Korea: A meta-frontier non-radial directional distance functionanalysis. *Energy Policy, 56*, 653–662. doi: https://doi.org/10.1016/j.enpol.2013.01.033
Zhu, J. (2001). Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research, 129*(2), 443–455. doi: Doi: 10.1016/s0377-2217(99)00433-6
Zhu, J. (2009). *Quantitative Models for Performance Evaluation and Benchmarking: DEA with Spreadsheets* (2nd ed.). Boston: Springer.
Zhu, Q., Aparicio, J., Li, F., Wu, J., & Kou, G. (2022). Determining closest targets on the extended facet production possibility set in data envelopment analysis: Modeling and computational aspects. *European Journal of Operational Research, 296*(3), 927–939. doi: https://doi.org/10.1016/j.ejor.2021.04.019
Zhu, Q., Wu, J., Ji, X., & Li, F. (2018). A simple MILP to determine closest targets in non-oriented DEA model satisfying strong monotonicity. *Omega, 79*, 1–8. doi: https://doi.org/10.1016/j.omega.2017.07.003
Zofio, J. L. (2007). Malmquist productivity index decompositions: a unifying framework. *Applied Economics, 39*(18), 2371–2387. doi: Doi 10.1080/00036840600606260
成刚, & 钱振华. (2012). 卫生体系效率评价的概念框架与测量方法——兼论应用数据包络分析的方法学问题. *中国卫生政策研究, 5*(3), 52–60.
马占新. (2002). 一种基于样本前沿面的综合评价方法. *内蒙古大学学报(自然科学版)*(6), 606–610.
马占新, & 马生昀. (2009). 基于 C^2W 模型的广义数据包络分析方法研究. *系统工程与电子技术*(2), 366–372.