数据包络分析软件 MaxDEA X

Data Envelopment Analysis Software MaxDEA X

2026-10-26

北京瑞沃迈迪软件有限公司

maxdea@maxdea.com

https://maxdea.com

目录

内容简介 **1**

第 1 章　MAXDEA 软件的基本操作 **2**

1.1　MaxDEA 软件的主要特点与支持的操作系统 2

1.1.1 主要特点 *2*

1.1.2 支持的操作系统 *2*

1.2　MaxDEA 软件的基本操作 3

1.3　数据格式 5

DMU 名称 6

面板数据格式 6

1.4　分析结果 7

第 2 章　DEA 基础模型 **8**

2.1 技术效率的概念 8

2.2 基于规模报酬不变的 CCR 模型 10

2.2.1 投入导向 CCR 模型的规划式 *10*

2.2.2 投入导向 CCR 模型得出的技术效率值的本质 *14*

2.2.3 产出导向 CCR 模型的规划式 *15*

2.2.4 产出导向 CCR 模型得出的技术效率值的本质 *16*

2.2.5 投入导向 CCR 模型图解 *17*

2.2.6 产出导向 CCR 模型图解 *18*

2.3 基于规模报酬可变的 BCC 模型 20

2.3.1 投入导向 BCC 模型的规划式 *20*

2.3.2 投入导向 BCC 模型得出的技术效率值的本质 *21*

2.3.3 产出导向 BCC 模型的规划式 *21*

2.3.4 产出导向 BCC 模型得出的技术效率值的本质 *22*

2.3.5 投入导向 BCC 模型图解 *22*

2.3.6 产出导向 BCC 模型图解 *23*

2.4 不同 DMU 之间效率值的可比性问题 23

2.5 投入导向和产出导向效率值的关系 25

2.6 规模效率问题 26

2.7 DEA 模型的命名问题 28

2.8 强有效、弱有效与松弛变量问题 29

2.9 比例改进与松弛改进的关系 33

2.10 DMU 数量、投入产出指标和模型导向的选择 34

2.10.1 DEA 模型对 DMU 数量的要求 *34*

2.10.2 投入和产出指标的选择 *34*

2.10.3 模型导向的选择 *36*

2.11 MAXDEA 求解 DEA 基础模型示例 37
2.11.1 模型设置 *38*
2.11.1.1 详细设置方式 38
2.11.1.2 快捷方式 39
2.11.2 分析结果 *39*
2.11.2.1 效率值与松弛变量 40
2.11.2.2 参考标杆与投影值（目标值） 41
2.11.2.3 投入和产出的权重系数 43
2.11.2.4 包络模型与乘数模型对偶解的关系 43
2.11.2.5 规模效率 43
2.12 对投影值容易产生的误解 44
2.13 包络模型与乘数模型输出结果之间的关系 45
2.13.1 包络模型与乘数模型投影图示 *45*
2.13.2 为什么乘数模型会存在多个最优解 *45*
2.13.3 为什么乘数模型结果中的权重会出现 0 ..... *46*

第 3 章 规模报酬 47

3.1 DEA 模型的规模报酬类型 47
3.2 DMU 规模报酬状态的判断 50
3.2.1 通过包络模型判断规模报酬状态 *50*
3.2.2 通过乘数模型判断规模报酬状态 *52*
3.3 MAXDEA 求解规模效率和规模报酬状态图示 56
3.4 规模报酬状态与模型导向的关系 56

第 4 章 DEA 模型的距离函数 57

4.1 径向距离 57
4.2 至前沿最远距离（SBM 模型） 59
4.2.1 SBM 模型 *59*
4.2.2 加权 SBM(ERM)模型 *62*
4.2.3 MSBM 模型 *62*
4.2.4 MaxDEA 求解 SBM(ERM)模型和加权 SBM(ERM)模型示例 *63*
4.3 至弱有效前沿最近距离 66
4.3.1 至弱有效前沿最近距离的计算方法 *66*
4.3.2 MaxDEA 求解 MinDW 模型示例 *67*
4.4 至强有效前沿最近距离 69
4.4.1 非导向至强有效前沿最近距离的计算方法 *69*
4.4.2 MaxDEA 求解非导向 MinDS 模型示例 *81*
4.4.2.1 至强有效前沿最小距离 CRS 模型 82
4.4.2.2 至强有效前沿最小距离 VRS 模型 83
4.4.3 MinDS 模型的特殊性 *85*
4.4.3.1 MinDS 模型 CRS 效率值可能会大于 VRS 效率值 85

4.4.3.2 扩展参考集后 MinDS 模型效率值有可能增大 86
4.4.3.3 单调性问题 88
4.4.3.4 投入和产出导向 MinDS 模型 88
4.4.4 MinDS 模型算法的新进展 91
4.5 方向距离函数 92
4.5.1 方向距离函数模型 92
4.5.2 方向距离函数模型的乘数形式 94
4.5.3 方向距离函数模型中对非期望产出的处理及对弱可处置性是否合理的讨论 95
4.5.4 方向距离函数模型的几种特例 103
4.5.5 一般方向距离函数模型效率值的计算方法 104
4.5.6 MaxDEA 求解方向距离函数模型示例 107
4.5.6.1 方向距离函数模型示例 107
4.5.6.2 包含非期望产出的方向距离函数模型示例 110
4.5.7 对包含非期望产出 DEA 模型的结果容易产生的误解 112
4.6 加权加性距离（包括 SBM 方向距离函数和非径向方向距离函数） 116
4.6.1 不加权的加性距离（简单加性距离） 116
4.6.2 加权加性距离 117
4.6.3 加权加性距离之简单加权的变形表达：基于方向距离函数之和的 SBM（slacks-based measure on the sum of directional distance functions） 119
4.6.4 加权加性距离之 RAM 模型 121
4.6.5 加权加性距离之 BAM 模型 122
4.6.6 加权加性距离之方向 SBM（directional slacks-based measure） 124
4.6.7 加权加性距离之非径向方向距离函数(non-radial directional distance function) 126
4.6.8 各类加权加性距离之间的关系 132
4.7 EBM 混合距离 134
4.7.1 EBM 模型 134
4.7.2 EBM 模型的缺陷与改进方法 136
4.7.3 MaxDEA 软件求解混合距离函数模型操作方法 143
4.8 包含价格信息的距离 144
4.8.1 成本效率、收益效率模型 144
4.8.2 MaxDEA 软件求解成本效率等模型的操作方法 145
4.9 单个投入产出指标无效率程度分析 146

第 5 章 超效率模型 147

5.1 径向超效率模型 147
5.2 方向距离函数超效率模型 150
5.3 SBM(ERM)超效率模型 151
5.4 加权加性距离超效率模型 157
5.5 超效率模型的投影分析 158
5.6 通过超效率模型探测异常数据 159

第 6 章　其他模型与方法 160

6.1 乘数比值约束模型 160
6.1.1 对两项投入的权重比值进行约束 160
6.1.2 对一项产出和一项投入的权重比值进行约束 161
6.1.3 对一项产出和另一项产出的权重比值进行约束 162
6.2 FDH 模型 162
6.3 包含非期望产出的模型 164
6.3.1 包含非期望产出的径向模型 164
6.3.2 包含非期望产出的 SBM 模型 164
6.3.3 包含非期望产出的方向距离函数模型 166
6.3.4 包含非期望产出的加权加性模型 166
6.3.4.1 包含非期望产出的简单加性模型 166
6.3.4.2 包含非期望产出的加权加性模型 166
6.3.5 通过负数指标实现对非期望产出的处理 166
6.4 MAXDEA 软件中对负数的处理 169
6.5 外部不可控因素或不可控投入/产出指标 169

第 7 章　广义 DEA 模型 171

7.1 广义 DEA 模型 171
7.2 群组参比模型 174
7.2.1 自我参比(Self-benchmarking) 174
7.2.2 交叉参比（Cross-benchmarking） 175
7.2.3 向下参比（Downward-benchmarking） 175
7.2.4 向上参比（Upward-benchmarking） 176
7.2.5 下方邻群参比（Lower-adjacent-benchmarking） 176
7.2.6 上方邻群参比（Upper-adjacent-benchmarking） 177
7.2.7 窗口参比(Window-benchmarking) 177
7.2.8 固定参比(Fixed-benchmarking) 179
7.3 群组参比模型的数据格式 180

第 8 章　面板数据模型 182

8.1 面板数据 182
8.2 生产率变化、技术变化与技术效率变化三者之间的关系 187
8.3 MALMQUIST 指数的基本原理 189
8.4 各类 MALMQUIST 指数计算方法 193
8.4.1 不同时期 MI 的计算参比不同时期的前沿 193
8.4.1.1 相邻参比（双指数平均） 194
8.4.1.2 序列参比（双指数平均） 196
8.4.1.3 窗口参比（双指数平均） 199
8.4.1.4 对 TC 的进一步分解 200
8.4.1.5 跨期参比引起的 VRS 模型无可行解问题 202

8.4.1.6 双期前沿（单一指数） 202
8.4.1.7 双期前沿对 TC 的进一步分解 203
8.4.1.8 MI 不可累乘问题 204
8.4.2 所有时期 MI 的计算均参比同一个前沿 204
8.4.2.1 固定参比（单一指数） 205
8.4.2.2 全局参比（单一指数） 206
8.4.2.3 固定窗口参比 208
8.4.2.4 同一前沿对 TC 的进一步分解 208
8.5 各类 MI 比较 210
8.6 文献中常用的 MALMQUIST 指数分解方法 214
8.6.1 R Färe et al. (1992)的分解方法 214
8.6.2 R. Färe et al. (1994)的分解方法 214
8.6.3 Ray & Desli (1997)的分解方法 215
8.6.4 Zofio (2007)的分解方法 215
8.7 LUENBERGER 指数 216
8.8 MALMQUIST-LUENBERGER 指数 217
8.9 COST-MALMQUIST 指数 217
8.10 HICKS-MOORSTEEN TFP 指数(HMB 指数) 218
8.11 对面板数据分析的几点建议 218
8.12 面板数据的窗口 DEA 模型 219
8.13 在不能获得 TFP 的情况下，如何分析 TFP 的影响因素 219
第 9 章 新型两阶段网络 DEA 与广义网络 DEA 221
符号及数据说明 222
9.1 两阶段网络 DEA (规模报酬不变 CRS) 223
9.1.1 两阶段网络 DEA 模型 223
9.1.2 基于中间流虚拟价值守恒的两阶段网络 DEA 模型 224
9.1.3 两阶段网络 DEA 模型(CRS)的效率分解 226
9.1.4 中间流柔性协同机制 226
9.1.5 为什么两阶段网络模型的效率分解要使用乘积形式 227
9.2 两阶段网络模型（规模报酬可变 VRS） 228
9.2.1 模型定义 228
9.2.2 效率分解 229
9.3 广义网络 DEA 模型 230
9.3.1 模型定义 230
9.3.2 模型求解 230
9.3.3 系统效率的计算方法 231
9.3.4 与前面的两阶段模型的兼容性 232
9.4 基于全维有效面(FDEF)的广义网络 DEA 233
9.4.1 全维有效面(FDEF)简介 233
9.4.2 基于 FDEF 仿射包的广义网络 DEA 模型 235

9.5 网络 DEA 的软件操作演示 236
9.5.1 定义数据 *236*
9.5.2 设置 DEA 模型的常规选项 *236*
9.5.3 设置其他模型选项 *236*
9.5.4 定义网络结构 *237*
9.5.5 输出结果 *237*
附录：CRS 两阶段网络 DEA 效率乘积分解证明 238

内容简介

数据包络分析（DEA）是一种应用非常广泛的效率分析方法，本书通过方法介绍与软件操作相结合，数学模型与图示相结合，旨在为广大 DEA 方法的应用者提供一本容易看懂、简单实用的指导书。

本书第 1 章介绍了 MaxDEA 软件的基本操作，第 2 章介绍了 DEA 基础模型（CCR 和 BCC），第 3 章介绍了 DEA 模型的规模报酬类型与规模弹性的计算方法，第 4 章介绍了 DEA 模型的各种距离函数（包括非期望产出），第 5 章介绍了超效率模型（包括非期望产出）并给出了 VRS 超效率模型无可行解问题的解决方案，第 6 章简要介绍其他 DEA 模型，第 7 章是广义 DEA 模型，第 8 章是面板数据模型。第 9 章是 MaxDEA 团队开发的新型两阶段网络模型和广义网络模型。在各章中均包含了 MaxDEA 软件求解示例或操作方法。

第1章 MaxDEA 软件的基本操作

1.1 MaxDEA 软件的主要特点与支持的操作系统

1.1.1 主要特点

1) 使用简单，原生跨平台支持 Windows、macOS 和 Linux。
2) 支持大型数据 DEA 模型，求解速度极快。
3) 全面支持 MinDS 模型最新进展：强单调性算法，CRS、VRS、NIRS 和 NDRS，非导向、投入导向和产出导向。
4) MaxDEA 最重要的特点是包含了大量 DEA 方法，并尽可能提供各类 DEA 方法之间的组合应用。只需同时设置多个模型选项即可建立组合模型。例如成本模型与 Malmquist 模型组合为成本 Malmquist 模型，非期望产出模型与方向距离函数模型、Malmquist 模型组合为包含非期望产出的方向距离函数 Malmquist 模型（这一组合通常称为 Malmquist-Luenberger 指数）等等。

1.1.2 支持的操作系统

1) Windows 7（64 位）或更高版本。
2) macOS 10.11 或更高版本。
3) Linux 各发行版本(x86-64)。

1.2 MaxDEA 软件的基本操作

MaxDEA 提供了简单明了的导航栏：第 1 步 - 第 2 步 - 第 3 步。



图 1-1 MaxDEA 软件的导航栏

第1步是新建一个空白的MaxDEA模型文件(文件名可以使用中文)。



图 1-2 新建模型图示

第2步是导入和定义数据。



图 1-3 导入数据图示



图 1-4 定义数据图示

第3步是设置和运行模型。



图 1-5 设置和运行模型图示

1.3 数据格式

MaxDEA 支持的数据类型包括：

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

   文本文件需要通过扩展名来表示数据格式：

   *.csv: 逗号(,)分隔的文本文件(最常用的文本数据格式)

   *.tsv: 制表符分隔的文本文件

   *.ssv: 分号(;)分隔的文本文件

   *.psv: 竖线(|)分隔的文本文件

   扩展名为.txt和.zip的文本文件，用户可以在导入数据时指定格式。扩展名为.zip的文件是压缩的.txt或.csv文件。
4) Stata (*.dta)
5) SPSS (*.sav; *.zsav)
6) SAS (*.sas7bdat; *.xpt; *.v8xpt)
7) dBase, FoxPro (*.dbf)
8) Access 数据库
9) SQLite 数据库(*.sdb)

如果数据文件是电子表格 (Excel, Open Office Calc 或文本文件)，要求数据的第一行是变量名称(支持使用中文)，并且只能有一行表示变量的名称。MaxDEA 对变量的名称及其排列顺序没有特殊要求。投入（产出）变量的变量名称可以自由设定。

表 1-1 正确的数据格式

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

表 1-2 错误的数据格式

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

当数据量很大时，往往存在错误或异常数据。在导入数据之前，请使用合适的数据处理软件(例如统计分析软件)对数据进行清洗。Excel 只是电子表格，没有严格的数据类型，容易存在难以察觉的错误，因此 Excel 并不适合做数据清洗工作。MaxDEA X 支持投入和产出变量数值为 0。如果在投入和产出变量中有空值，MaxDEA X 自动将空值视同为 0。

DMU 名称

在数据表中必须有一列用于表示各个 DMU 的名称(DMU Name)。DMU 名称可以是文字（支持中文），也可以是数字或字母数字组合。

面板数据格式

在面板数据中必须有一列数据来表示数据的时间。MaxDEA 对面板数据的排列顺序没有特殊要求。面板数据可以按照 DMU 的名称排序，也可以按照时间排序，也**可以不排序**。

表示时间的变量必须是**整数**，可以是连续的整数序列，例如

1, 2, 3……

2001, 2002, 2003……

可以是非连续的整数序列，例如

1, 2, 5, 8……

2001, 2005, 2009……

200101, 200102, 200302……

MaxDEA 允许使用非平衡面板数据，即允许部分时期的数据缺失（整条数据缺失）。缺失数据对应的时期，其分析结果也相应缺失。

表 1-3 面板数据格式示例（平衡面板）

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

表 1-4 面板数据格式示例（非平衡面板）

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 分析结果

分析结果保存在与模型文件相同位置的一个单独的文件夹内，文件夹的名称为当前日期。如果该文件夹已经存在，则使用“当前日期+数字后缀”。软件运行完毕会显示结果保存的位置。

第2章 DEA 基础模型

2.1 技术效率的概念

技术效率指一个生产单元的生产过程达到该行业技术水平的程度。技术效率反映的是一个生产单元技术水平的高低，因此称为“技术”效率。技术效率可以从投入和产出两个角度来衡量，在投入既定的情况下，技术效率由产出最大化的程度来衡量；在产出既定的情况下，技术效率由投入最小化的程度来衡量(Farrell, 1957)。

当不考虑生产规模的影响时，技术效率可通过产出/投入的比值来定量地测量，即表达为 y/x。如果不考虑生产规模的影响，并且不涉及多个时期时，技术效率与生产率所表达的含义是相同的。当生产过程仅涉及一种投入和一种产出时，可以计算各生产单元的产出/投入比值，即每消耗一个单位的投入所生产的产品数量，来反映各生产单元技术效率的高低。如果将各单元的产出/投入比值除以其中的最大比值，就可以将产出/投入比值标准化为 0-1 之间的数值，这样可以更好地反映被评价单元与最优单元之间技术效率的差距（表 2-1）。

表 2-1 用产出/投入比值测量技术效率

单位	x(投入)	y(产出)	y/x	y/x(标准化)
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

上述方法虽然简单，但仅适用于单投入、单产出的情况。如果生产过程涉及的投入和产出不止一项，则无法直接计算单一的比值。在多投入多产出的情况下，有两种思路来实现计算产出投入比值。

一是对各投入和产出指标赋予一定的权重，然后计算加权产出/加权投入的比值，可作为反映技术效率的指数。这是 DEA 乘数模型的思路。

二是以一组投入产出标杆数据（x*, y*）作为参照，被评价的这组投入产出数据（x_k, y_k）与之做比较，与参考标杆的比值作为被评价数据的技术效率，表示为$E_{(x_k,y_k)}$：$E_{(x^*,y^*)}=\frac{y_k/y^*}{x_k/x^*}$ 。这是 DEA 包络模型的思路。

上述两种方法构建的线性规划模型互为对偶关系，得出的技术效率值相等。

我们从第一种方法出发，来理解 DEA 的基本原理。

假设有 m 种投入和 q 种产出，则加权投入表示为

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

加权产出表示为

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

接下来的问题是如何确定反映各项投入和产出之间相对重要程度的权重系数。一种方法是采用固定的权重，例如通过专家咨询或研讨等主观的形式确定各项指标的权重；另一种方法是通过数据本身获得投入和产出的权重，DEA 乘数模型就是采用这种方法。

2.2 基于规模报酬不变的 CCR 模型

1978 年 Charnes，Cooper 和 Rhodes 三人在《欧洲运筹学杂志》(European Journal of Operational Research)上发表了论文《Measuring the efficiency of decision making units》，创立了 DEA 理论方法(A Charnes, Cooper, & Rhodes, 1978)。在后来的 DEA 文献中，以 Charnes，Cooper 和 Rhodes 三人姓氏的首字母来命名他们创立的第一个 DEA 模型，即 CCR 模型。CCR 模型假设规模报酬不变（constant returns to scale，CRS），其得出的技术效率包含了规模效率的成分，因此通常被称为综合技术效率。

DEA 将效率的测度对象称为决策单元（decision making unit，DMU），DMU 可以是任何具有可测量的投入、产出（或输入、输出）的部门、单位，例如厂商、学校、医院、项目执行单位（区域），也可以是个人，例如教师、学生、医生等。DMU 之间须具有可比性。

2.2.1 投入导向 CCR 模型的规划式

上节中谈到可以通过计算产出投入比来测量技术效率，现在接着这个话题来阐述 DEA 的基本原理。假设我们要测量一组共 n 个 DMU 的技术效率，记为 DMU_j (j=1,2…,n)；每个 DMU 有 m 种投入，记为 x_i (i=1,2…,m)，投入的权重表示为 v_i (i=1,2…,m)；q 种产出，记为 y_r (r=1,2…,q)，产出的权重表示为 u_r (r=1,2…,q)。当前要测量的 DMU 记为 DMU_k，其产出投入比表示为

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

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

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

接下来给要测量的技术效率值附加一项条件，将所有 DMU 采用上述权重得出的效率值 θ_j 限定在[0, 1]的区间内，即

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

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

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

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

Charnes，Cooper 和 Rhodes 三人创立的第一个 DEA 模型是基于规模报酬不变假设的，该模型表示为

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

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

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

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

这一非线性规划模型的含义在于，在所有 DMU 的效率值都不超过 1 的条件下，使被评价 DMU 的效率值最大化。因此模型确定的权重 u 和 v 是对被评价 DMU_k 最有利的，从这个意义上讲，CCR 模型是对被评价 DMU 的无效率状况做出的一种保守的估计，因为它采用的权重是最有利于被评价者的，采用其它任何权重得出的效率值都不会超过这组权重得出的效率值。

为什么说 CCR 模型是基于规模报酬不变的呢？假设一项生产技术的规模报酬不变，则在技术效率保持不变的条件下，如果一个 DMU 的投入变为原来的 t 倍(t>0)，其产出也会相应变为原来的 t 倍。反过来说，假设被评价 DMU_k 的投入和产出都变为原来的 t 倍，在规模报酬不变假设下，其技术效率应保持不变。那么 CCR 模型在这种情况下得出的技术效率是否是保持不变的呢？

DMU_k 的投入和产出都变为原来的 t 倍后，CCR 模型的目标函数为

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

由于 t 同时存在于分子和分母，约掉后与原来的目标函数等价。同理，与 DMU_k 对应的约束也与原来的约束等价。因此，当被评价 DMU_k 的投入和产出都增加为原来的 t 倍时，CCR 模型与原模型等价，得出的效率值不变，与规模报酬不变的假设相符。

CCR 模型(2.5)存在的问题是它是非线性规划并且存在无穷多个最优解。假设向量 u^*和 v^*是模型(2.5)的一个最优解[i]，则 tu^*和 tv^*肯定也是模型的最优解(t>0)。

由于 $\sum_{i=1}^{m} v_k x_{ij} > 0$，模型(2.5)的约束等价于

[i] 在本书中，采用在变量加上标*的方式表示线性规划中该变量的解。

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

令 $t=\dfrac{1}{\sum_{i=1}^{m}v_i x_{ik}}$，则

模型(2.5)的目标函数变为

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

再令 $\mu=tu$，$v=tv$[i]，

非线性模型(2.5)变换为等价的线性规划模型

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

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

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

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

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

模型(2.9)是以求解 DMU k 为例来表述投入导向 CCR 模型的线性规划，对于每个 DMU 都要分别建立规划式。

表 2-2 线性规划式示例数据

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

以表 2-2 中的数据为例，需分别求解 4 个线性规划模型，从 A 到 D 依次为

DMU A

[i] 注意等号前面使用的是希腊字母 μ 和 ν，等号后面使用的是英文字母 u 和 v，采用不同的字母是因为变换后的模型或之前的模型是两个不同的线性规划，虽然其目标函数最优解相等。

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

模型(2.9)的对偶模型为[i]

[i] 原模型与对偶模型之间的转换关系请参阅相关教材，例如 Hillier, F. S., & Lieberman, G. J. (2010). Introduction to Operations Research: McGraw-Hill Higher Education.

$$\min \theta$$

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

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

$$\lambda \ge 0$$

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

对偶模型(2.10)中，λ 表示 DMU 的线性组合系数，模型的最优解 θ^*代表效率值，θ^*的范围为(0, 1]。

（$x=\sum_{j=1}^{n}\lambda_j x_j$， $y=\sum_{j=1}^{n}\lambda_j y_j$）可以看作是一个虚拟的 DMU，其投入不高于 DMU_k 的投入，产出不低于DMU_k的产出。如果DMU_k处于无效率的状态，则最优解构建的虚拟DMU（$\hat{x}=\sum_{j=1}^{n}\lambda_j^* x_j$，$\hat{y}=\sum_{j=1}^{n}\lambda_j^* y_j$）就是被评价 DMU_k 的投影值。

模型的目标函数最优解为 θ^*，1-θ^*表示在当前技术水平下，被评价 DMU_k 在不降低产出水平的条件下，其投入能够缩减的最大程度。θ^*越小，表示投入可以缩减的幅度越大，效率越低。$\theta^*=1$ 时说明被评价 DMU 位于前沿面上，在不减少产出的条件下，其各项投入没有等比例下降的空间，处于技术有效的状态；$\theta^*<1$ 说明被评价 DMU 为技术无效率状态，在不减少产出的条件下，其各项投入可以等比例下降的比例为（1-θ^*）。

在模型(2.9)中，投入和产出权重系数与投入和产出在形式上是乘数和被乘数的关系，模型(2.9)通常称为 DEA 的乘数形式(multiplier form)。其对偶模型(2.10)确定的前沿形似包络，将所有 DMU 包裹在内（详见后面的图解），通常称为 DEA 的包络形式(envelopment form)。

CCR 对偶模型(2.10)是以产出既定的条件下，各项投入可以等比例缩减的程度来对无效率的状况进行测量，因此被称为投入导向（input-oriented）的 CCR 模型。

2.2.2 投入导向 CCR 模型得出的技术效率值的本质

前面提到多投入多产出情况下技术效率的测量有两种思路，第二种思路是采用参比标杆的方法。

模型(2.10)可以等价表示为

$$\min \theta$$

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

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

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

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

在此模型中，我们可以把参比标杆理解为$\left(x^*, y^*\right) = (\sum_{j=1}^{n}\lambda_j x_{ij} + s_i^-, \sum_{j=1}^{n}\lambda_j y_{rj} - s_r^+)$，则被评价 DMU 的产出投入比值与参比标杆的产出投入比值之比（即二者的比值比）为

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

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

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

上述推导过程有利于帮助我们理解 CCR 模型得出的技术效率值的本质，即投入导向 CCR 模型得出的技术效率值 θ^*是被评价 DMU 的生产率与参比标杆的生产率的比值。

2.2.3 产出导向 CCR 模型的规划式

产出导向 CCR 模型的规划式为

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

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

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

$$v \geq 0; \mu \geq 0$$

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

其对偶模型为

$$\max \phi$$

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

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

$$\lambda \ge 0$$

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

对偶模型(2.14)是以投入既定的条件下，各项产出可以等比例增长的程度来对无效率的状况进行测量，因此被称为产出导向（output-oriented）的 CCR 模型。

模型的最优解为 φ^*。在当前技术水平下，被评价 DMU_k 在不增加投入的条件下，其产出能够增长的最大比例为 φ^*-1。φ^*越大，表示产出可以增长的幅度越大，效率越低。由于 $\varphi^*\ge 1$，所以一般采用 $1/\varphi^*$表示效率值[i]。

2.2.4 产出导向 CCR 模型得出的技术效率值的本质

模型(2.14)可以等价地表示为

$$\max \phi$$

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

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

$$\lambda, s^-, s^+ \ge 0 \qquad (2.15)$$

在此模型中，我们同样可以把参比标杆理解为$\left(x^*, y^*\right)=(\sum_{j=1}^{n}\lambda_j x_{ij}+s_i^-, \sum_{j=1}^{n}\lambda_j y_{rj}-s_r^+)$，则被评价 DMU 的产出投入比值与参比标杆的产出投入比值之比（即二者的比值比）为

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

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

[i] 在 MaxDEA 软件的分析结果中，产出导向模型的效率值（Efficiency Score）＝$1/\varphi^*$。

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

从上述推导过程可以看出，产出导向 CCR 模型得出的技术效率值 $1/\varphi^*$同样是被评价 DMU 的生产率与参比标杆的生产率的比值。

2.2.5 投入导向 CCR 模型图解

我们下面以图示的方式来展示投入导向 CCR 模型的基本原理。假设有 7 个 DMU（表 2-3），均有 2 种投入（x_1、x_2）和 1 种产出（y）。以单位产出消耗的投入 x_1 的数量为横坐标(x_1/y)，以单位产出消耗的投入 x_2 的数量为纵坐标(x_2/y)，各 DMU 单位产出的投入情况可用图 2-1 来表示。

表 2-3 投入导向 CCR 模型示例数据

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

在图 2-1 中， DMU 可用坐标系内的点表示。从被评价 DMU 的坐标点，分别向两个坐标轴作垂线，在两条垂线与坐标轴围成的区域内（含边界），任意点的坐标值均小于或等于被评价 DMU 的坐标值。以 G 为例，阴影区域 GQOP 内所有点的坐标值，记为点 M(m_1, m_2)，均小于或等于 G 的坐标，记为 G(g_1, g_2)，即 $m_1 <= g_1$, $m_2 <= g_2$。处于该区域内所表示的 DMU 与 G 相比，其每生产一个单位的产出所消耗的两种投入的数量均小于或等于 G 的消耗（仅当 M 与 G 重合时，两者相等），这表示，在该区域内除 G 自身外，所有点的技术效率均高于 G，从图中可以看出，在 7 个 DMU 中，B、C、E 三个点与 G 相比，效率更高。

在 7 个 DMU 中，有 4 个 DMU（A、B、C、D），在其垂线与坐标轴围成的区域内，均没有包含任何其它的 DMU 或其它 DMU 的线性组合，这表示 A、B、C、D 处于技术效率的前沿上，四点连接构成的曲线及其延长线称为有效前沿，位于前沿的 DMU 其效率为 1，被前沿包裹的 DMU 效率在 0～1 之间。从形状上看，前沿凸向原点，形似包络，包裹着所有 DMU。在多维空间中，投入导向 CCR 模型的前沿为凸向原点的凸多面体的顶面。

这就是为什么将这一分析方法称为数据包络分析（Data Envelopment Analysis，DEA）的原因所在。

以 G 为例，G(g_1, g_2)与坐标原点 O(0, 0)的连线和前沿曲线的交点记为 G'(g'_1, g'_2)，G'称为 G 在前沿上的投影（projection）。 G 与处于前沿曲线上的 G'点相比，其每生产一个单位的产出所多消耗的

两种投入的数量分别为 $g_1 - g'_1$ 和 $g_2 - g'_2$，多消耗的比例分别为$(g_1 - g'_1) / g_1$ 和$(g_2 - g'_2) / g_2$，有效消耗的比例分别为 $1 - (g_1 - g'_1) / g_1 = g'_1 / g_1$ 和 $1 - (g_2 - g'_2) / g_2 = g'_2 / g_2$。上述两个比例用坐标内的线段来表示分别为 RG'/PG 和 SG'/QG，根据几何定理可知 RG'/PG＝SG'/QG＝OG'/OG。G 的效率值可表示为 OG'/OG。G 的效率部分体现为 OG'，无效率部分体现为 GG'。

从 DEA 基本原理的图示可以看出，DEA 测度的效率为相对效率，是被评价 DMU 相对于“领先”DMU 的效率，DEA 得出的效率值是被评价 DMU 与处于前沿上的投影点相对比较得出的。

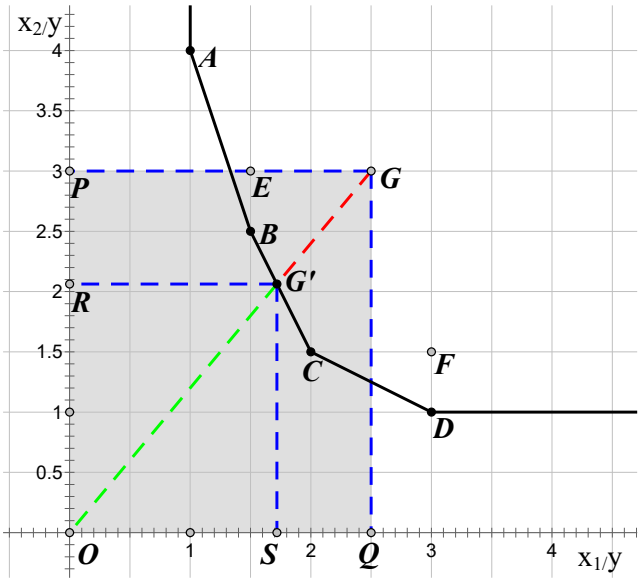


图 2-1 投入导向 CCR 模型基本原理图示

2.2.6 产出导向 CCR 模型图解

假设有 7 个 DMU，均有 1 种投入和 2 种产出（y_1 和 y_2）。以单位投入所生产的产出 y_1 的数量为横坐标(y_1/x)，以单位投入所生产的产出 y_2 的数量为纵坐标(y_2/x)，各 DMU 单位投入的产出情况可用图 2-2 来表示。

表 2-4 产出导向 CCR 模型示例数据

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

以 G 为例，坐标原点 O 与 G 连线的延长线与前沿曲线的交点记为 G'，G'称为 G 在前沿上的投影（projection）。每消耗一个单位的投入，G'点产量与 G 产量的比值分别为 RG'/PG 和 SG'/QG，根

据几何定理可知 RG'/PG＝SG'/QG＝OG'/OG。G 的效率状况可表示为 φ＝OG'/OG。由于 φ 的取值范围为[1, ∞)，在实际应用中一般采用 φ 的倒数作为效率值，θ=1/φ＝OG / OG'，θ 取值范围为(0, 1]。

在多维空间中，产出导向 CCR 模型的前沿为凹向原点的凸多面体的顶面。

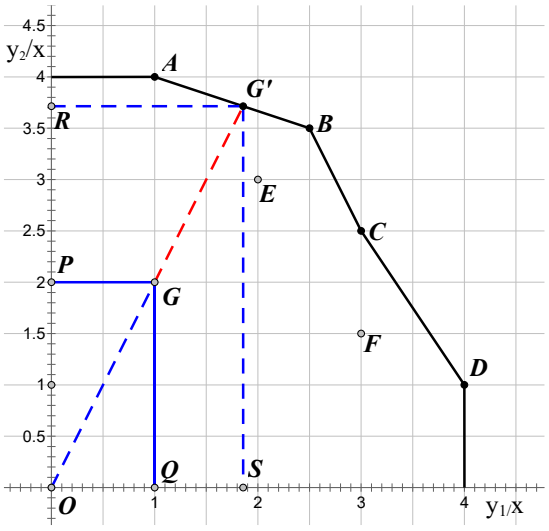


图 2-2 产出导向 CCR 模型基本原理图示

2.3 基于规模报酬可变的 BCC 模型

CCR 模型假设生产技术的规模报酬不变，或者虽然生产技术规模报酬可变，但假设所有被评价 DMU 均处于最优生产规模阶段，即处于规模报酬不变阶段。但实际生产中，许多生产单位并没有处于最优规模的生产状态，因此 CCR 模型得出的技术效率包含了规模效率的成分。1984 年 Banker，Charnes 和 Cooper 三人在《Management Science》杂志上发表了《Some models for estimating technical and scale inefficiencies in data envelopment analysis》，提出了估计规模效率的 DEA 模型(Banker, Charnes, & Cooper, 1984)。这一方法的提出对于 DEA 理论方法具有重要的意义，在以后的文献中将此模型称为 BCC 模型（以三位作者的姓氏首字母命名）。BCC 模型基于规模报酬可变（variable returns to scale, VRS），得出的技术效率排除了规模的影响，因此称为“纯技术效率”（pure technical efficiency，PTE）。

2.3.1 投入导向 BCC 模型的规划式

BCC 模型是在 CCR 对偶模型(2.10)的基础上增加了凸性约束 $\sum_{j=1}^{n}\lambda_j=1$，其作用是基于规模报酬可变技术，构建去除规模效率之后的纯技术效率前沿，从而使被评价 DMU 与其投影点之间的比较是基于同等规模报酬（详见后文图解）。

$$\min \theta$$

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

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

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

$$\lambda \ge 0$$

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

BCC 模型(2.17)的对偶规划式为[i]

[i] μ_0是自由变量，取值范围为（－∞， ＋∞），因此在线性规划式中μ_0前面的符号是正号还是负号，不影响目标函数的结果。与原文献(Banker et al., 1984)保持一致，本书中采用负号。

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

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

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

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

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

2.3.2 投入导向 BCC 模型得出的技术效率值的本质

投入导向 BCC 模型得出的技术效率值 θ *也是被评价 DMU 的生产率与参比标杆的生产率的比值。与投入导向 CCR 模型的不同之处在于 BCC 模型的参考标杆位于纯技术效率前沿上。

2.3.3 产出导向 BCC 模型的规划式

产出导向 BCC 模型的规划式为

$$\max \phi$$

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

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

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

$$\lambda \ge 0$$

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

产出导向 BCC 模型是在产出导向 CCR 模型(2.14)的基础上增加了凸性约束 $\sum_{j=1}^{n} \lambda_j = 1$。

其对偶规划式为

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

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

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

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

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

2.3.4 产出导向 BCC 模型得出的技术效率值的本质

产出导向BCC 模型得出的技术效率值 1/φ*同样是被评价 DMU 的生产率与参比标杆的生产率的比值。与产出导向 CCR 模型的不同之处在于 BCC 模型的参考标杆位于纯技术效率前沿上。

2.3.5 投入导向 BCC 模型图解

下面以图示的方式来展示投入导向 BCC 模型的基本原理。假设有 4 个 DMU（表 2-5），均有 1 种投入（x）和 1 种产出（y）。以投入 x 为横坐标，以产出 y 为纵坐标，各 DMU 的投入产出情况如图所示(图 2-3)。

表 2-5 投入导向 BCC 模型示例数据

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

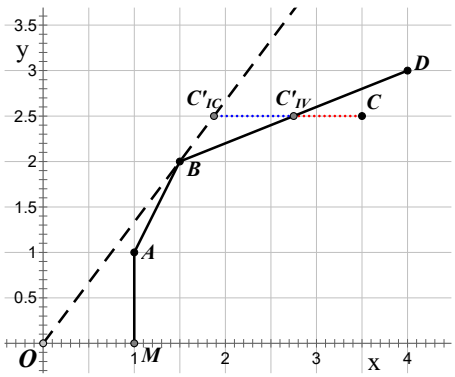


图 2-3 投入导向 BCC 模型基本原理图示

假设规模报酬不变（CRS），生产前沿可用 OB 射线表示，B 是唯一有效 DMU。假设规模报酬

可变（VRS），则生产前沿为 MABD 构成的凸向左侧的曲线。在 CRS 模型中，B 点的产出投入比最高，A、C、D 三点与 B 点相比，均处于无效率状态，其中 A 点的投影处于 OB 上，λ 值小于 1；C、D 两点的投影处于 OB 延长线上，λ 值大于 1。而在 VRS 模型中，受到凸性约束$\sum_{j=1}^{n}\lambda_j=1$的限制，A、C、D 三点均无法向 OB 射线作投影，因为它们在 OB 射线上的投影点要么$\sum_{j=1}^{n}\lambda_j<1$，要么$\sum_{j=1}^{n}\lambda_j>1$，无法满足$\sum_{j=1}^{n}\lambda_j=1$的条件。凸性约束$\sum_{j=1}^{n}\lambda_j=1$使得生产前沿退缩，并使 A、D 两点处于前沿上，ABD 及 A 点垂直向下的延长线构成了“纯技术效率”生产前沿曲线。VRS 模型中 C 点为无效 DMU，其投入无效率体现为 CC'_{IV}，而在 CRS 模型中其投入无效率体现为 CC'_{IC}。

2.3.6 产出导向 BCC 模型图解

下面再以图示的方式来展示产出导向 BCC 模型的基本原理。同样以表 2-5 的数据为例，在 CRS 模型中，产出导向与投入导向的前沿相同，也是 OB 射线。但是在 VRS 模型中，产出导向的前沿与投入导向的前沿并不完全相同，产出导向的“纯技术效率”生产前沿曲线为 ABD 及 D 点平行于 x 轴的延长线（图 2-4）。C 点在前沿上的投影则完全不同，其无效率体现为 CC'_{OV}，而在 CRS 模型中其产出无效率体现为 CC'_{OC}。

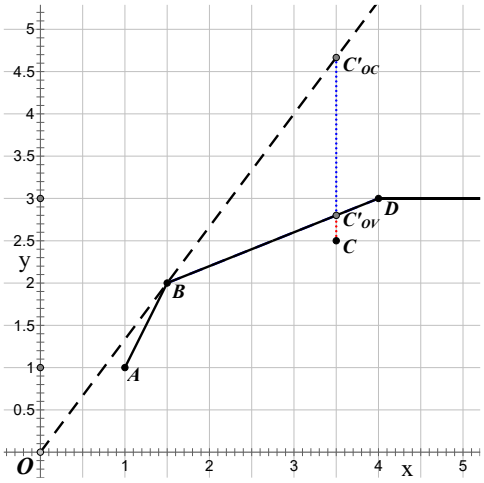


图 2-4 产出导向 BCC 模型基本原理图示

2.4 不同 DMU 之间效率值的可比性问题

DEA 模型得出的技术效率值是被评价 DMU 的生产率与参比标杆的生产率的比值。如果所有 DMU 都参考同一标杆，则无疑各个 DMU 的效率值相互之间具有可比性。但实际上，不同 DMU 的参考标杆是不同的。

不同 DMU 之间效率值仍然具有可比性是基于生产前沿在技术效率上的等效性，即生产前沿上的所有点（每个点代表一个虚拟的 DMU），其技术效率均为 100%。虽然不同 DMU 的参考标杆不同，但所有参考标杆的技术效率等同。

生产前沿是 DEA 模型的核心，对效率值的测量发挥着关键作用。稳定的生产前沿是得出可靠结果的保障。构建稳定的生产前沿需要足够数量的 DMU。

2.5 投入导向和产出导向效率值的关系

在 CRS 模型中，投入导向和产出导向得出的效率值相等，这一点可通过图示的方式来说明（图 2-5）。C 点的效率值，在投入导向模型中（θ^*）可表示为 QC'_{IC}/QC，在产出导向模型中（$1/\varphi^*$）可表示为 NC/NC'_{OC}。通过三角几何定理很容易证明 $C'_{IC}C/QC=CC'_{OC}/NC'_{OC}$，然后即可得出 $QC'_{IC}/QC=NC/NC'_{OC}$。但是在 VRS 模型中，不存在这样的关系，即 VRS 模型中投入导向和产出导向的效率值不等。

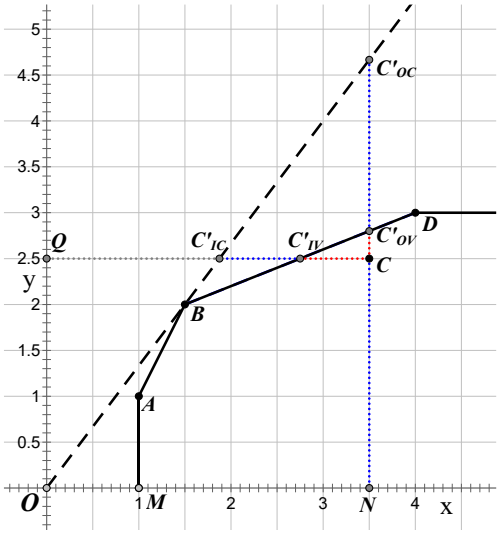


图 2-5 CRS 模型中投入导向与产出导向的效率值相等图解

2.6 规模效率问题

BCC 模型的提出是为了求解 VRS 生产技术下 DMU 的技术效率，但同时 BCC 模型也为计算规模效率提供了方法。如果生产技术是规模报酬可变的，采用 CRS 模型得出的效率值（technical efficiency, TE）就并非是纯粹的技术效率，而是包含了规模效率的成分，这使得求解规模效率成为可能。对 VRS 生产技术而言，既然 VRS 模型得出的效率值才是技术效率（称为“纯技术效率”，pure technical efficiency， PTE），那么通过比较计算 CRS 效率值和 VRS 效率值就可以分离出规模效率值 (scale efficiency， SE)，计算方法为 SE=TE/PTE。

在图 2-6 中投入导向的规模无效率体现为 $C'_{IV}C'_{IC}$，在图 2-7 中产出导向的规模无效率体现为 $C'_{OC}C'_{OV}$。

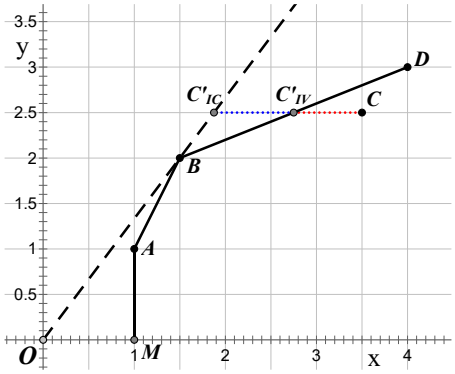


图 2-6 投入导向模型规模效率图示

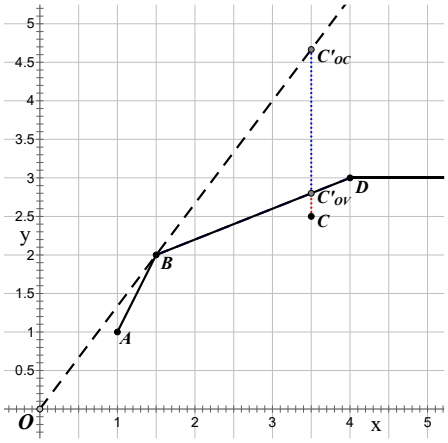


图 2-7 产出导向模型规模效率图示

从以上图示可以看出，DEA 模型得出的一个 DMU 的规模效率并非该 DMU 自身的规模效率，而是通过该 DMU 计算得出的该 DMU 的前沿投影点的规模效率。投入导向和产出导向模型的前沿投影点不同，因此得出的规模效率也不同。

需要注意的一点是，当生产技术为规模报酬可变（VRS）时才存在规模效率问题。在求解规模效率时，只是借用 CRS 模型（假设所研究的 DMU 处于 CRS 状态）来分离出规模效率。需明确其中的逻辑关系。

根据规模效率的定义，计算规模效率需采用径向距离。当采用其他距离类型时[i]，得出的 TE/PTE 比值并不是“规模效率”(Scale Efficiency)，在 MaxDEA 软件中将这一类似比值称为“规模效应”(Scale Effect)，以示区别。

William W Cooper, Seiford, & Tone (2007)将 SBM 效率值与径向效率值之间的差异称为 MIX (Mix Efficiency，混合效率)，即 SBM 效率值 = 径向效率值×MIX。按照这样的定义，可得出以下关系

$$\text{SBM规模效应}=\frac{\text{SBM CRS效率值}}{\text{SBM VRS效率值}}=\frac{\text{径向 CRS效率值}\times\text{CRS MIX}}{\text{径向 VRS效率值}\times\text{VRS MIX}}$$

$$=\text{规模效率}\times\text{Scale MIX}$$

。

[i] 其他距离类型详见第 4 章。

2.7 DEA 模型的命名问题

CCR 模型标志着 DEA 方法的诞生，而 BCC 模型的提出对于 DEA 方法的发展也具有非常重要的意义。因此，这两个模型采用作者姓氏命名得到了 DEA 领域的一致认可。但是，随着 DEA 理论方法的快速发展，大量新的 DEA 方法不断出现，如果继续采用作者姓氏的方式进行命名，容易引起混乱。DEA 模型宜根据模型的主要特征命名。根据这样的命名方法，CCR 模型为 CRS 径向 DEA 模型；BCC 模型为 VRS 径向 DEA 模型。径向(radial)的含义是指无效率的测量方式为，投入能够等比例减少的程度，或产出能够等比例增加的程度。为便于介绍更多的 DEA 模型，本书后面的章节中将不再使用 CCR 和 BCC 两个名称。

2.8 强有效、弱有效与松弛变量问题

为什么在 DEA 模型中有松弛变量呢？在径向 DEA 包络模型的线性规划式(2.10)中(2.14)，约束条件是以不等式的形式来表示的，而非等式。这可以看作是一种“松”的约束，这是松弛变量存在的基础[i]。

可以用生活中的以下问题来帮助理解径向 DEA 模型中的松弛变量：有多条长短不一的绳子，现在将所有绳子的一端固定在墙上，然后将所有绳子的另一端对齐后往外拉，当拉不动时，说明最短的一根绳子已经拉紧了，不能继续拉了。但是，这时其它绳子可能还是松的，如果放开已经拉紧的绳子，松弛的绳子还可以继续拉动，每根松弛的绳子可以继续拉动的距离就是各个指标的松弛变量值。

还可以从包络模型的线性空间中去理解松弛问题产生的原因：DEA 模型的前沿是由分段线性函数构成的，而分段线性函数在空间坐标系中会出现与坐标轴平行的情况，这是松弛问题产生的根源。

首先看一下单投入单产出的例子，还以表 2-5 中的数据为例。在 CRS 模型中，单投入单产出的前沿为 OB 射线，不可能出现与坐标轴平行的情况，因此单投入单产出的 CRS 径向模型松弛变量肯定为 0，即不存在松弛问题。但是，VRS 前沿会出现与坐标轴平行的区段，如果某个 DMU 的投影落在平行区段内，则会存在松弛问题。例如在表 2-5 数据的基础上增加一个 S 点（x=2, y=0.5）。在投入导向 VRS 径向模型中，S 点在前沿的投影点为 S'，但是 S'点与前沿上的 A 点相比，S'存在产出不足（即 S'A 段），存在松弛问题（图 2-8）。

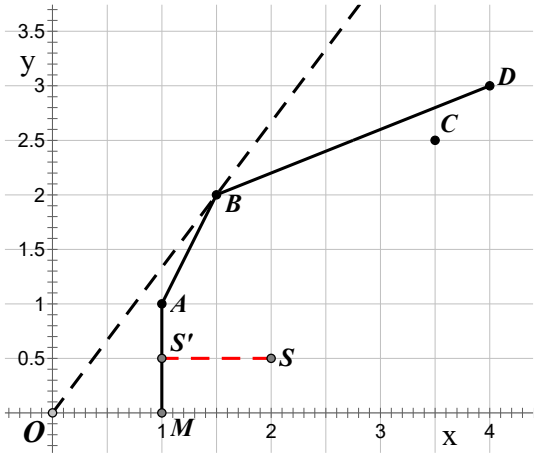


图 2-8 单投入单产出情况下 VRS 模型中的松弛问题图解

在多投入多产出的情况下，无论是 CRS 模型还是 VRS 模型都会出现松弛问题。

[i] 在生产可能集理论中，DEA 模型的生产可能集表示为

$$\left\{(x,y): x \geq \sum_{j=1}^{n}\lambda_j x_{ij},\ y \leq \sum_{j=1}^{n}\lambda_j y_{rj}; i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\right\}$$

这种采用不等号的“松”的约束代表着投入和产出的可自由处置性（free disposability），其含义为：如果投入 x 能生产产出 y，则更多的投入 x+Δx 也可以生产 y；如果 x 能生产 y，则 x 也可以生产更少的产出 y-Δy。

图 2-9 显示了在两项投入和一项产出的投入导向 CRS 模型中，存在松弛变量的问题。E 点在前沿上的投影落在与坐标轴平行的区段内的 E'上，其单位产出的投入 x_2/y 存在的松弛为 E'B。

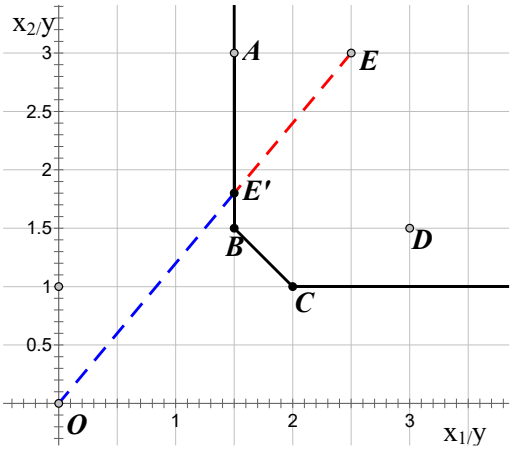


图 2-9 两投入一产出情况下 CRS 模型中的松弛问题图解

在 DEA 模型中，当 DMU 数量过少时，构建的前沿面会不够精细，在被评价 DMU 的投影区域内，更容易出现与坐标轴平行的区段，从而出现松弛问题。

求解松弛变量值可以有多种方法。以投入导向的模型为例，如果只是在模型的约束中增加松弛变量，在目标函数中不增加松弛变量，即

$$\min \theta$$

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

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

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

这种方法通常称为一阶段（one stage）方法。在多数情况下，一阶段方法一般能获得正确的松弛变量值，例如图 2-8 的例子。但一阶段方法并不能保证一定能获得完全的松弛变量值，例如在图 2-9 中，在 E 的投影点所处的平行于坐标轴的区段内存在两个 DMU，这时，采用一阶段方法无法求解出投入 x2/y 的松弛变量。在求解 DEA 模型时，如果用户只关心效率值，可以选择一阶段方法。

为了解决一阶段方法获得的松弛变量值不完全的问题，需要在模型的目标函数中增加投入和产出的松弛变量 s^-和 s^+。以投入导向径向模型为例，以模型(2.21)为基础，在目标函数中增加松弛变量后的规划式为：

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

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

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

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

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

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

ε 是一个常量，表示非阿基米德无穷小，可理解为一个足够小的正数，例如在实际应用中设置为 0.00001。ε 设置不当会产生计算误差，ε 数值设置过大时，甚至会造成目标函数无边界。

模型(2.22)的乘数形式(对偶模型)为

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

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

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

$$v\geq\varepsilon$$

$$\mu\geq\varepsilon$$

$$\mu_0\ free$$

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

ε 数值设置过大时会造成模型无可行解[i]。

为了消除包络模型(2.22)中 ε 产生的误差，可采用两阶段（two stage）方法求解投入和产出松弛变量(Ali & Seiford, 1993)。

在第一阶段，求解模型(2.17)；然后在第二阶段，求解以下规划式：

[i] 根据对偶理论，原模型如果目标函数无边界，则其对偶模型无可行解。

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

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

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

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

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

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

θ^*表示在第一阶段取得的 θ 的最优解。

在图 2-8 中，产出 y 的松弛变量为 0.5。在图 2-9 中，投入 x_2/y 的松弛变量为 0.3。

在产出导向模型中，也可以通过类似方式，求解松弛问题。

通过 DEA 模型可以得出被评价 DMU 的效率状态：有效（率）还是无效（率）。DEA 有效分为两种情况：

一是强有效（strong efficiency），它是指生产处于这样的一种状态，任何一项投入的数量都无法减少，除非减少产出的数量或增加另外一种投入的数量；任何一项产出的数量都无法增加，除非增加投入的数量或减少另外一种产出的数量。这种生产状态是一种 Pareto 最优状态。在径向 DEA 模型中，判断是否为强有效的标准是 $\theta^*=1$，并且 s^{-*}和 s^{+*}均为 0。

另外一种是弱有效（weak efficiency），它是指无法等比例减少各项投入的数量，除非减少产出的数量；无法等比例增加各项产出的数量，除非增加投入的数量。在这种生产状态下，虽然不能等比例减少投入或增加产出，但是某一项或几项（但不是全部）投入可能可以减少；或者某一项或几项（但不是全部）产出可能可以增加，因此这样的生产状态被称为弱有效。在径向 DEA 模型中，判断是否弱有效的标准是 θ^*是否等于 1，不需要考虑是否存在松弛问题。

是否有效和是否松弛是两个不同的问题。在径向 DEA 模型中，并非一定存在松弛问题，也可能出现所有松弛变量结果均为 0 的情况。

第一种情况是只有一种投入和一种产出的 CRS 模型。单投入单产出 CRS 模型的前沿为一个点，不存在与坐标轴平行的问题，所以也就不存在松弛问题。

第二种情况是所有无效 DMU 在前沿上的投影点均位于前沿上的强有效部分，在这种情况下，所有无效 DMU 的所有松弛变量结果也会全部为 0，不出现松弛问题。

2.9 比例改进与松弛改进的关系

无效 DMU 在前沿上的投影点代表其目标值，如果投入的改进值用负数表示，产出的改进值用正数表示[i]，则被评价 DMU 的投入和产出投影值的计算方法可统一表示为：

投影值＝原始值+改进值。

无效率 DMU（简称为无效 DMU）的改进值包括两部分，一是比例改进值(Proportionate Movement)[ii]，二是松弛改进值(Slack Movement)。即

投影值＝原始值+比例改进值+松弛改进值。

如果无效 DMU 仅完成比例改进，则该 DMU 改进后有可能是弱有效：如果存在松弛问题，则为弱有效；如果不存在松弛问题，则为强有效。只有完成比例改进和松弛改进后，才能保证无效 DMU 变为强有效。

以图 2-10 为例，D 为无效 DMU，当完成比例改进后，由 D 点转移到前沿上的 C 点，由于 C 点处于前沿的强有效部分，不存在松弛问题，所以 D 在完成比例改进后即为强有效。E 也是无效 DMU，当完成比例改进后，由 E 点转移到前沿上的 E'点，由于 E'点处于前沿的弱有效部分（与坐标轴平行的部分），存在松弛问题，所以 E 在完成比例改进后仅为弱有效；在此基础上，再完成松弛改进后，由 E'点转移到 B 点，这时才为强有效。

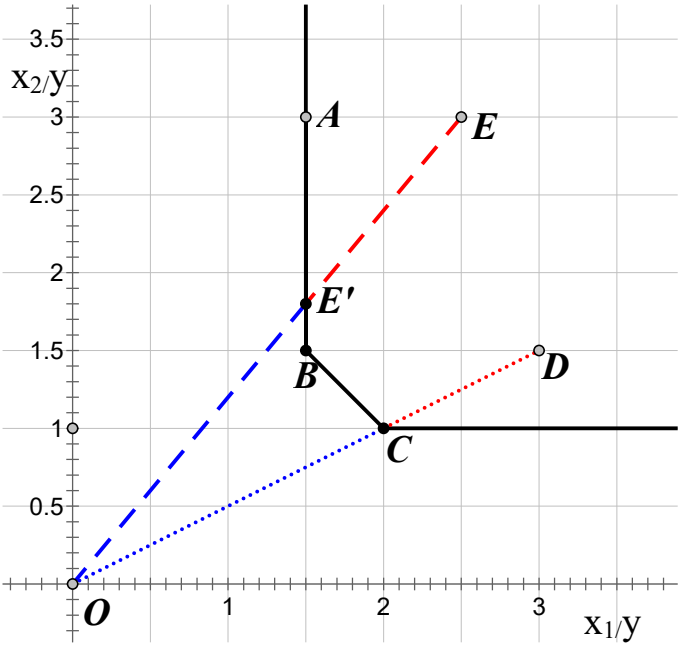


图 2-10 两投入一产出情况下 CRS 模型中的松弛问题图解

[i] 注意在超效率模型中，超效率 DMU 的“改进值”的符号会呈现与标准效率模型相反的情况，“改进值”在超效率模型中具有不同的含义，详见“第 5 章 超效率模型”。

[ii] 在径向模型中，比例改进值一般称为径向改进值（Radial Movement），表示各项投入（或产出）等比例改进的数量。但在方向距离函数模型中的，这一部分改进并非等比例改进，而是按照方向向量的方向改进。为统一这部分改进值的名称，在 MaxDEA 软件的分析结果中，将其统称为比例改进值。

2.10 DMU 数量、投入产出指标和模型导向的选择

2.10.1 DEA 模型对 DMU 数量的要求

DEA 为非参数前沿分析方法，虽然相对于参数方法， DEA 对 DMU 数量的要求相对较少，但是，如果 DMU 数量过少，例如 DMU 的数量比投入产出指标的数量还要少（n<m+q），则很容易出现大部分甚至全部 DMU 均有效的结果，使 DEA 失去对 DMU 效率进行区分的能力。一般来说，DMU 的数量不应少于投入和产出指标数量的乘积，同时不少于投入和产出指标数量的 3 倍(William W Cooper et al., 2007)，即

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

但这只是一个粗略的指导性原则，具体要根据 DEA 分析结果来判断。在实际应用中，往往 DMU 数量是固定的，当模型区分能力不足时，只能通过减少投入或产出指标数量来提高区分度。

当 DMU 数量固定时，纳入模型的投入产出指标数量越多，得出的效率值越大。在原来模型基础上（无论是 CCR 还是 BCC 模型），如果增加一项投入或产出指标，新模型得出的效率值一定是大于或等于原模型得出的效率值。

2.10.2 投入和产出指标的选择

从生产可能集的角度考虑，DEA 模型的投入指标(x)和产出指标(y)应大致满足以下关系：

1）x 能生产 y；

2）y 是由 x 生产出来的。

在实际应用中，应注意区分投入产出指标与效率（或生产率）影响因素的不同。以医院为例，其主要产出指标是门急诊和住院服务，主要投入指标包括卫生技术人员、床位及其它医疗设备设施等。当前正在进行的医药卫生体制改革会对医院的效率（或生产率）产生影响，但是医改是效率（或生产率）的外部影响因素，并非投入指标。

DEA 不要求投入指标或产出指标之间不存在高度相关性（共线性），共线性的存在不会导致错误的分析结果。实际上，在许多生产实践中，往往客观上就要求两种（或以上）生产要素要按一定比例投入。以医院为例，医生和护士的比例就存在着一个合理的范围，在实际数据中，医生的数量和护士的数量存在高度相关性。

但是，在指标数量较多而造成模型区分能力不足时，从尽量减少投入产出指标数量并尽可能包含更多的生产要素的角度出发，可首先考虑从模型中排除高度相关的指标。

DEA 模型是基于生产可能集理论的线性规划方法，生产可能集是 DMU 的线性组合，从理论上讲，DEA 模型中的投入和产出指标必须可以线性相加。率（或比值）是由分子指标和分母指标相除得出的，如果各 DMU 的率指标的分母数值不同（实际应用中，往往是不同的），就会产生错误的生产可能集，还可能会产生不合逻辑的结果(Emrouznejad & Amin, 2009)。例如，反映卫生体系投入和产出的指标有很多是率或比值的形式，如卫生总费用占 GDP 的比例、服务覆盖率、死亡率、发病

率、患病率等。如果直接采用率作为投入或产出指标，在 CRS 模型中，投影分析的结果中可能会出现率的投影值大于 100%的情况。虽然采用 VRS 模型可以避免出现率的投影值大于 100%的问题(Hollingsworth & Smith, 2003)，但是 VRS 模型仍然存在 DEA 模型确定的生产前沿处于生产可能集之外的问题(Emrouznejad & Amin, 2009)。

在 VRS 模型中，当率指标的分母数值在各 DMU 之间差别不大时，可近似认为该率指标满足可线性相加的条件；或者由于不满足可线性相加而造成的误差为研究目的所允许时，该率指标可以作为投入或产出指标。

Emrouznejad & Amin (2009)建议，当投入或产出指标中存在率时，可以采用两种方法对这些指标进行处理。

以产出导向的 VRS 模型为例，第一种方法是将率的分子作为产出指标，而将率的分母作为投入指标，其线性规划为

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

其中，N 和 D 分别表示率指标的分子和分母，n_0 和 d_0 分别表示被评价 DMU 该率指标的分子和分母。

第二种方法是在线性规划式的左边将率指标拆分为分子指标和分母指标，分别进行线性运算，但是在方程的右边不对被评价 DMU 的率指标进行拆分，其规划式为

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

其中，N 和 D 分别表示率指标的分子和分母，r_0 表示被评价 DMU 的率指标。

第一种方法的优点是保持了模型的线性规划，缺点是增加了投入或产出指标的数量，从而增加

了对样本量的要求；第二种方法的优点是没有增加投入和产出指标数量，但模型变为非线性规划问题。这两种方法的前提都是必须获得率指标的分子和分母数据。

2.10.3 模型导向的选择

按照对效率的测量方式，DEA 模型可分为投入导向（input-oriented）、产出导向(output-oriented)和非导向（non-oriented，后面介绍）。投入导向模型是从投入的角度对被评价 DMU 无效率程度进行测量，关注的是在不减少产出的条件下，要达到技术有效各项投入应该减少的程度；产出导向模型是从产出的角度对被评价 DMU 无效率的程度进行测量，关注的是在不增加投入的条件下，要达到技术有效各项产出应该增加的程度；非导向模型则是同时从投入和产出两个方面进行测量。

具体到本章中介绍的径向模型，投入导向是指在产出既定的条件下，用各项投入可以**等比例**缩减的程度来对无效率的状况进行测量；而产出导向则是在投入既定的条件下，各项产出可以**等比例**增长的程度来对无效率的状况进行测量。在 DEA 模型中，“在产出既定的条件下”实际意义是指 “在产出不减少的条件下”；而“在投入既定的条件下”实际意义是指“在不增加投入的条件下”。

模型导向的选择主要取决于分析目的。如果分析目的只是获得各单位的效率值，上述三种导向均可。如果需要做进一步的投影分析，从管理角度考虑，如果把减少投入作为对无效率单位提高效率的主要途径，应选择投入导向模型；如果把增加产出作为提高效率的主要途径，则应选择产出导向模型。导向的选择要结合分析的具体研究领域去理解。例如在卫生领域，如果在卫生资源投入不足的背景下，选择投入导向模型会使得投影分析的结果不容易进行解释。因为在投入导向模型中，根据投影分析的结果，无效率的单位要达到有效率的状态，其改进目标是减少投入，这似乎与卫生资源投入不足的背景产生矛盾（但实质上并不矛盾），使得分析结果容易产生误解。另外，在分析卫生服务效率时，在需求不足的背景下（当需求是决定卫生服务产出数量的首要因素时），选择产出导向模型，会使得投影分析所确定的产出目标客观上难以实现，从而失去实际指导意义(成刚 & 钱振华, 2012)。

2.11 MaxDEA 求解 DEA 基础模型示例

本节利用实际数据来说明 MaxDEA 软件求解基础 DEA 模型的过程与结果。表 2-6 是 2011 年各省(市、自治区)医院的部分投入和产出指标。以床位数和卫生技术人员数作为投入指标，以诊疗人次数和入院人数作为产出指标，建立 DEA 模型。DEA 基础模型包括投入导向 CRS 模型、产出导向 CRS 模型、投入导向 VRS 模型和产出导向 VRS 模型 4 种径向 DEA 模型。为节省篇幅，本节仅对 CRS 模型进行详细演示。

表 2-6 各地区医院部分投入产出指标（2011 年）

地区	床位数	卫技人员数	诊疗人次数	入院人数
安徽	140997	132739	63344221	4391516
北京	87596	128644	104340626	1882593
福建	89947	93898	72059496	3087618
甘肃	66661	53127	27719369	1710826
广东	246050	289388	293784406	7993062
广西	95752	105773	64449855	3258438
贵州	78368	69320	29105344	2433370
海南	21367	26329	12692336	611621
河北	187504	183683	82489933	5675067
河南	239793	231149	114076405	7043628
黑龙江	129449	127358	47962073	2992970
湖北	152062	147628	84006816	4805124
湖南	168428	152904	63238363	5393968
吉林	94636	86278	38087139	2259264
江苏	221674	214938	166944454	6433837
江西	87184	93287	46383376	3000964
辽宁	171032	156893	70270696	4032159
内蒙古	72871	74177	31379522	1747905
宁夏	22037	21830	12373335	602227
青海	18586	16689	8918596	469811
山东	280385	281654	134719233	8354234
山西	110741	116660	37390228	2332714
陕西	114339	123537	52692308	3236715
上海	87548	95198	113669022	2282773

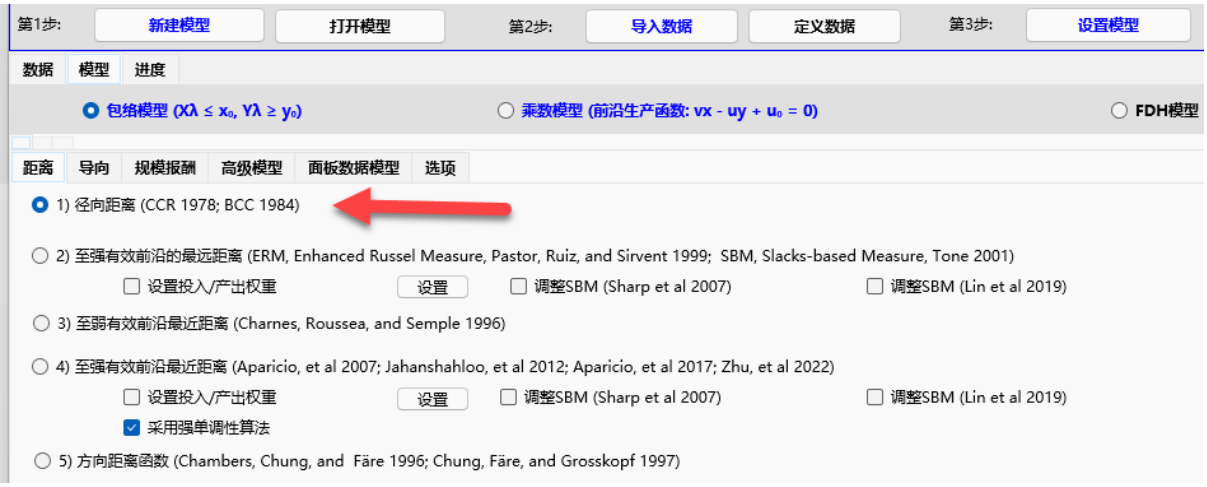
四川	211524	194752	110475915	6489314
天津	40787	53543	52164779	1037054
西藏	6314	5843	3998564	120766
新疆	97436	84202	37383273	3115413
云南	126318	90330	62327579	3736495
浙江	162905	187137	181663463	4899310
重庆	74827	66653	38726559	2116631

2.11.1 模型设置

2.11.1.1 详细设置方式

径向模型的基本设置包括“距离”（距离函数类型）、“导向”、“规模报酬”三个方面。

距离选择“1) 径向距离 (CCR 1978; BCC 1984)”。



导向选择“1) 投入导向”或者 “2) 产出导向”。



规模报酬：CCR 模型选择“不变(CRS)”，BCC 模型选择“可变(VRS)”。



2.11.1.2 快捷方式

针对 4 种基础 DEA 模型，MaxDEA 提供了快捷菜单。通过菜单“模型”，“一键运行基础模型”，可以直接运行 4 种基础 DEA 模型：投入导向 CRS 模型、产出导向 CRS 模型、投入导向 VRS 模型和产出导向 VRS 模型。注意通过快捷方式运行的是包络模型。

2.11.2 分析结果

包络模型的结果包括：

1） Score：效率值

2） Rank：效率值排序

3） Benchmark and Lambda：参考标杆及其系数

4） Proportionate Movement：等比例变化值（径向变动值），正数表示投影方向为增加数量，负数表示投影方向为减少数量

5） Slack Movement：松弛变量值，正数表示投影方向为增加数量，负数表示投影方向为减少数量

6） Projection：目标值（投影值）

7） Dual Price：投入和产出的影子价格，对应乘数模型分析结果中的投入和产出的权重[i]。对偶解的具体含义详见线性规划或运筹学教材。

乘数模型的结果包括：

1） Score：效率值

2） Rank：效率值排序

3） Weight：投入和产出的权重系数

[i] 在包络模型的对偶解中，投入的影子价格表示为负数，产出的影子价格表示为正数；在乘数模型中，投入和产出的权重系数均表示为正数。由于可能存在多个最优解，包络模型的投入和产出的影子价格与乘数模型的投入和产出的权重有可能不一致。

4） RTS Constant：规模报酬的常数项[i]

5） Benchmark and Lambda：参考标杆及其系数

6） Projection：目标值（投影值）

2.11.2.1 效率值与松弛变量

下表是投入导向 CRS 和产出导向 CRS 模型的效率值与松弛变量。在分析结果中，第一列 NO 是序号，第二列是 DMU 的名称，第三列是 Score，即效率值，在投入导向模型中，Score = θ^*，在产出导向模型中 Score = $1/\varphi^*$。松弛变量的结果表示形式为 Slack Movement (指标名称)，例如床位数的松弛变量结果表示为 Slack Movement (床位数)。对于无效 DMU 来说，其改进方向是减少投入，或增加产出。因此，在 MaxDEA 的分析结果中，投入的松弛改进值用负数表示，产出的松弛改进值用正数表示。

注意 DEA 模型中的 Slack（松弛变量）本身是**非负数**，在 MaxDEA 软件的输出结果中采用 Slack Movement 的名称并在其数值前面增加正负号，是用以说明投入产出松弛变化的方向（负数表示减少，正数表示增加）。这一点在后面第 5 章的超效率模型结果中尤为重要。如果需要模型中 Slack 的原始数值，取 Slack Movement 的绝对值即是。

表 2-7 投入导向 CRS 和产出导向 CRS 模型的效率值与松弛变量[ii]

	投入导向 CRS					产出导向 CRS				
DMU	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**
安徽	0.944	0	0	11623213	0	0.944	0	0	12309544	0
北京	0.917	0	-30637	0	212841	0.917	0	-33394	0	231997
福建	1.000	0	0	0	0	1.000	0	0	0	0

效率值与松弛变量的规律性及本例结果的特点如下：

1） 前文已指出，投入导向 CRS 和产出导向 CRS 模型的效率值相同，表中结果用实例验证了这一点。

2） 效率值等于 1 为（弱）有效，如果同时所有松弛变量均为 0，则为强有效。本例分析结果中，有 6 个 DMU 的效率值为 1，并且这 6 个 DMU 的所有投入和产出的松弛变量均为 0，说明 6 个 DMU 均为强有效。

3） 在投入导向的径向模型中，至少有一项投入的松弛变量为 0；在产出导向的径向模型中，

[i] CRS 模型此项为 0，因此在 CRS 模型的结果中不需要显示此项。

[ii] 为了减少表格的宽度，松弛变量的结果没有保留小数部分。在下一表格中作了相同处理。为减少篇幅，本书示例仅列出部分结果。

至少有一项产出的松弛变量为 0。以投入导向的模型为例，假设某 DMU 的结果中，所有投入的松弛变量都不为 0，说明仍然存在等比例改进的空间，所得到的解并非最优解，因此不存在所有投入的松弛变量都不为 0 的情况。在本例分析结果中，投入导向模型中，不存在床位数和卫技人员数的松弛变量均不为 0 的情况；产出导向模型中也不存在诊疗人次数和入院人数的松弛变量均不为 0 的情况。

2.11.2.2 参考标杆与投影值（目标值）[i]

表 2-8 列出了部分 DMU 的参考标杆和投影值。在分析结果中，参考标杆用 Benchmark(Lambda)表示，Lambda 是被参考的 DMU 的系数，有多个参考 DMU 时，用分号进行分隔，例如安徽的参考标杆为“福建(0.6362); 新疆(0.7791)”，这表示安徽在前沿上的投影点由福建和新疆的投入产出线性组合构成，福建和新疆的权重系数分别为 0.6362 和 0.7791。

表 2-8 投入导向 CRS 和产出导向 CRS 模型的参考标杆与投影值

	投入导向 CRS					产出导向 CRS				
DMU	参考标杆	床位数	卫技人员数	诊疗人次数	入院人数	参考标杆	床位数	卫技人员数	诊疗人次数	入院人数
安徽	福建(0.6362); 新疆(0.7791)	133136	125338	74967434	4391516	福建(0.6737); 新疆(0.8251)	140997	132739	79394136	4650828
北京	上海(0.9179)	80363	87385	104340626	2095434	上海(1.0005)	87596	95250	113731343	2284025
福建	福建(1.0000)	89947	93898	72059496	3087618	福建(1.0000)	89947	93898	72059496	3087618

无效 DMU 在前沿上的投影值代表其改进的目标值，假设其投入和产出指标达到了投影值，如果在当前的数据集中再次对其进行评价（其它 DMU 的投入和产出保持不变），则评价结果将变为效率值等于 1（弱有效或强有效）。

在 MaxDEA 的分析结果中，投入的改进值用负数表示，产出的改进值用正数表示。因此，被评价 DMU 的投入和产出投影值的计算方法可统一表示为：

投影值 ＝ 原始值 + 改进值[ii]。

改进值包括两部分，一是比例改进值(Proportionate Movement)[iii]，二是松弛改进值(Slack

[i] 在方便表述，在本书中，将最优解中 $\lambda^*>0$ 的 DMU 称为被评价 DMU 的参考标杆。参考标杆的线性组合构成被评价 DMU 的投影值（目标值）。在参考集中没有作为参考标杆的 DMU，其 $\lambda^*=0$。

[ii] 有的文献中，将投入的原始值和目标值之间的差距称为“投入冗余”。“冗余”是一个容易产生误导的名称。按照字面理解，冗余（redundancy）简单的说就是多余，投入冗余就是投入中多余的部分，有它没它结果都一样。在 DEA 的投影分析中，投入的原始值和目标值之间的差距代表的是被评价 DMU 和前沿之间的差距。与前沿相比由于被评价 DMU 技术效率低而多消耗的投入要素。这些多消耗的投入要素，对于被评价 DMU 来说，并非多余。如果单纯的让被评价 DMU 砍掉这些投入，其产出数量就会相应减少。减少或消除与前沿的差距，途径是改进技术和管理水平从而提高技术效率。

[iii] 在径向模型中，比例改进值一般称为径向改进值（Radial Movement），表示各项投入（或产出）等比例改进的数量。但在方向距离函数模型中的，这一部分改进并非等比例改进，而是按照方向向量的方向改进。为统一这部

Movement)[i]。即

投影值＝原始值+比例改进值+松弛改进值。

在投入导向径向模型中，投入的比例改进值＝-原始值×（1-θ*），产出的比例改进值为 0；在产出导向径向模型中，投入的比例改进值为 0，产出的比例改进值＝原始值×（φ*-1）。

表 2-9 以安徽为例，演示了通过改进值的方法计算投影值。

表 2-9 投入和产出投影值计算方法示例：改进值法（安徽）

	投入导向 CRS				产出导向 CRS			
	床位数	卫技人员数	诊疗人次数	入院人数	床位数	卫技人员数	诊疗人次数	入院人数
原始值	140997	132739	63344221	4391516	140997	132739	63344221	4391516
比例改进值	-7861	-7401	0	0	0	0	3740371	259312
松弛改进	0	0	11623213	0	0	0	12309544	0
投影值	133136	125338	74967434	4391516	140997	132739	79394136	4650828

此外，被评价 DMU 的投入和产出的投影值也可以通过参考标杆及其权重系数来计算：

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

下表同样以安徽为例，演示了通过标杆值法计算强有效投影值。由于原始值最大为 8 位数字，所以权重系数需要保留 9 位小数，以避免计算误差。两种方法计算的投影值相同。

表 2-10 投入和产出投影值计算方法示例：标杆值法（安徽）

	投入导向 CRS				产出导向 CRS			
	床位数	卫技人员数	诊疗人次数	入院人数	床位数	卫技人员数	诊疗人次数	入院人数
福建原始值	89947	93898	72059496	3087618	89947	93898	72059496	3087618
福建权重	0.636154651				0.673718522			
新疆原始值	97436	84202	37383273	3115413	97436	84202	37383273	3115413
新疆权重	0.779130551				0.825136911			
投影值	133136	125338	74967434	4391516	140997	132739	79394136	4650828

分改进值的名称，在 MaxDEA 软件的分析结果中，将其统称为比例改进值。

[i]如果在后续的分析中需要利用结果中的投影值，建议在“选项”界面，“松弛变量计算”选择两阶段方法。否则，计算得出的弱有效投影值仍然有可能是弱有效，虽然概率很低。

2.11.2.3 投入和产出的权重系数

选择乘数模型(菜单 Step 2: Run Model-Run Multiplier Model)。分析结果显示在结果窗口的“Weights (Dual Values)”表格中，投入产出的权重用 Weight(指标名称)表示，例如 Weight(床位数)表示投入指标“床位数”的权重系数。

2.11.2.4 包络模型与乘数模型对偶解的关系

包络模型的解对应乘数模型的对偶解；乘数模型的解对应包络模型的对偶解。在包络模型的对偶解中，投入产出的对偶价格相当于乘数模型中的投入产出权重系数。由于乘数模型经常存在多个最优解，所以两者在数值上有可能不相同。

2.11.2.5 规模效率

在“规模报酬”选项中，如果选择“**效率或 TFP 指数分解 (包含规模效率 / 规模效应)**”，软件会先后运行 CRS 模型和 VRS 模型，并提供两个模型的效率值及根据两个模型的效率值计算的规模效率值（Scale Efficiency）或规模效应值（Scale Effect）[i]。包络模型和乘数模型得出的规模效率是相同的。

在 MaxDEA 的结果窗口中，显示的三个效率值分别是

1） Technical Efficiency Score(CRS)：表示 CRS 模型获得的技术效率值。为与 VRS 模型得出的技术效率值相区别，通常将 CRS 模型得出的技术效率（Technical Efficiency, TE）称为综合技术效率。

2） Pure Technical Efficiency Score(VRS)：表示 VRS 模型获得的技术效率值，通常称为纯技术效率(Pure Technical Efficiency Score, PTE)。

3） Scale Efficiency Score/ Scale Effect Score：规模效率值(Scale Efficiency, SE)或规模效应值(Scale Effect, SE)。

表 2-11 列出了部分省份的综合技术效率(TE)、纯技术效率(PTE)和规模效率(SE)。三者之间的关系为:

规模效率值(SE) = 综合技术效率(TE)/纯技术效率(PTE)

表 2-11 规模效率

	投入导向模型			产出导向模型		
DMU	综合技术效率	纯技术效率	规模效率	综合技术效率	纯技术效率	规模效率

[i] 根据文献中对规模效率的定义，计算“规模效率”（Scale Efficiency）应采用径向距离。当采用其他距离类型时，得出的结果在 MaxDEA 软件中称为“规模效应”（Scale Effect）。

	（TE）	（PTE）	（SE）	（TE）	（PTE）	（SE）
安徽	0.9442	0.9549	0.9889	0.9442	0.9568	0.9869
北京	0.9174	0.9185	0.9989	0.9174	0.9175	0.9999
福建	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
甘肃	0.8324	0.8538	0.9750	0.8324	0.8473	0.9825
广东	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
广西	0.9901	0.9961	0.9940	0.9901	0.9964	0.9937
贵州	0.9631	0.9708	0.9920	0.9631	0.9698	0.9931
海南	0.8331	0.9415	0.8849	0.8331	0.9323	0.8936
河北	0.9039	0.9391	0.9625	0.9039	0.9452	0.9563
河南	0.8826	0.9615	0.9180	0.8826	0.9698	0.9101
黑龙江	0.6893	0.6900	0.9990	0.6893	0.7020	0.9820
湖北	0.9469	0.9704	0.9759	0.9469	0.9724	0.9738

2.12 对投影值容易产生的误解

在 DEA 模型的结果中，投影值（目标值）是被评价的 DMU 在前沿上的投影，一个 DMU 的当前值与投影值之间的相对差距代表着该 DMU 各项投入产出的无效率程度。

投影值又被称为目标值，容易产生的一种误解是，只要将一个无效的 DMU 的投入产出数值改为投影值，则该 DMU 将会成为有效的 DMU。这是一种对结果的机械的理解。单纯从计算过程看，如果将一个无效 DMU 的各项投入产出数值修改为其投影值（假设其他 DMU 保持不变），重新计算效率值，则该 DMU 确实会变为有效。但是现实生产过程中，如果单纯地减少一个 DMU 的投入数量，则其产出数量必然相应减少；如果要单纯地提高一个 DMU 的产出数量，则必须相应地增加其投入数量。如果单纯根据投影结果，要求一个 DMU 减少其投入同时保持产出不变（甚至增加产出），或者增加其产出同时保持投入不变（甚至减少投入），这是不可能实现的。

当前值与投影值之间的差距是一个 DMU 无效率的结果，不是无效率的原因。DEA 分析本身是不可能给出无效率的原因的，因此也不可能直接给出改进效率的途径。一个 DMU 之所以无效率，可能是由其生产技术落后、管理水平低下等原因造成的。DEA 分析是给出 DMU 无效率的状况（效率值和与前沿的差距），不是给出无效率的原因，更不可能直接给出改进效率的途径。

DEA 分析结果可以间接指导无效的 DMU 改进效率。例如分析结果给出了无效 DMU 的参考标杆。无效的 DMU 可以对标杆 DMU 进行深入分析，找出自身与标杆之间在技术和管理方面的差距，确定自己改进效率的途径。一个无效的 DMU 通过改进技术和管理水平，其技术效率得到提高，该 DMU 与前沿之间的差距就会缩小或者消失。

总之，无效 DMU 与前沿的差距，是其技术无效率的结果，不是技术无效率的原因。投入减少

或产出增加是无效率 DMU 通过改进技术和管理水平，达到技术有效之后的预期结果。提高技术效率的途径是改进技术和管理水平，不是减少投入或增加产出。

2.13 包络模型与乘数模型输出结果之间的关系

2.13.1 包络模型与乘数模型投影图示

包络模型的输出结果除了效率值之外，还包括被评价 DMU 在前沿上的投影点（参考标杆和投影值）。以图 2-11 中的 E 点为例，E 的前沿投影点为 E'，在包络模型的结果中表示为 E' = 0.5 A + 0.5B。被评价 DMU 在前沿上的投影点表示为该前沿段 AB 的顶点（A 和 B）的线性组合。AB 之间的任意一点可表示为 $P = \lambda_1 A + \lambda_2 B$，$\lambda_1 + \lambda_2 = 1$， $\lambda_1 \geq 0$， $\lambda_2 \geq 0$。

乘数模型的输出结果除了效率值之外，还包括投入和产出的权重和自由变量(u*或 v*)。E 点的乘数模型结果为投入 x 的权重为 0.2，产出 y 的权重为 0.1，自由变量为-0.4。乘数模型给出的结果对应的是 E 点在前沿上的投影点 E'所在的超平面（即直线 AB）的方程，即

$-2x + y + 4 = 0$。方程两边同乘以 0.1，即为 $-0.2x + 0.1y + 0.4 = 0$。

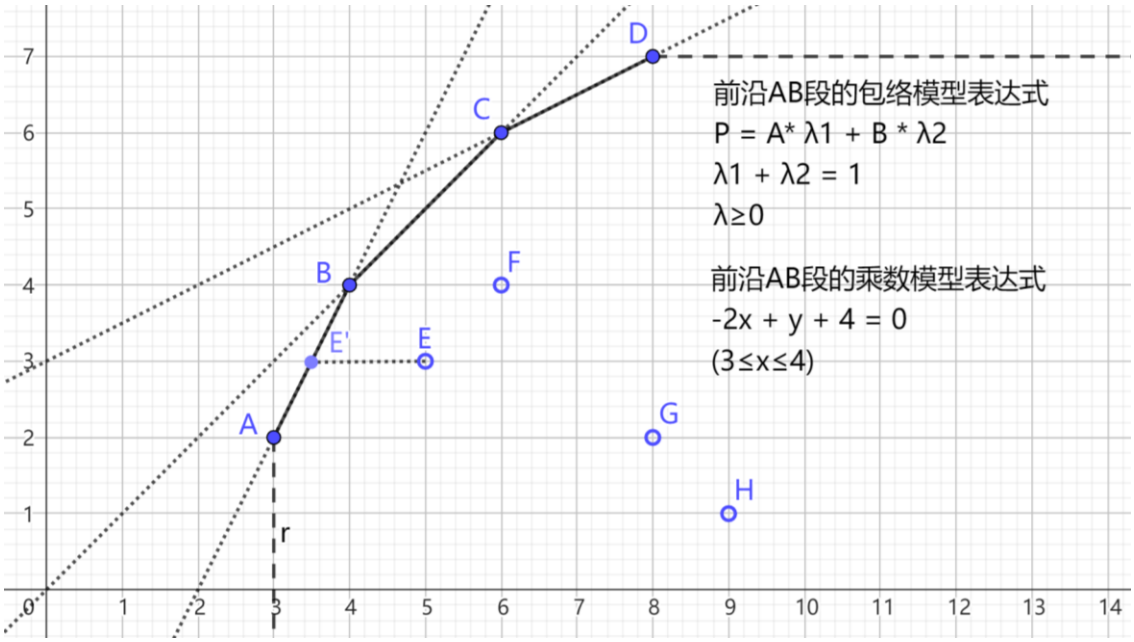


图 2-11 包络模型与乘数模型输出结果的关系图示(单投入单产出投入导向 VRS 模型)

2.13.2 为什么乘数模型会存在多个最优解

以图 2-11 中的 F 点为例，其投影点为 B，而 B 同时处于直线 AB 和直线 BC 两个超平面上，因此 F 存在两个最优解。G 的投影点为 A，也存在两个最优解。另外，前沿上的顶点（A、B、C、D）均同时处于两个超平面上，因此均存在多个最优解。

投入和产出指标数量越多，存在多个最优解的情况越多，且最优解的数量越多。例如在三投入三产出的 VRS 乘数模型中，最优解的数量可能会多达数百个（前沿上一个顶点可能会同时处于数百个超平面上）。

2.13.3 为什么乘数模型结果中的权重会出现 0

以图 2-11 中的 H 点为例，其投影点落在 A 点与 X 轴的垂线 r 上，而垂线 r 的方程为 -x + 0y + 3 = 0。y 的系数为 0，因此对应的产出权重为 0。当投影点所处的前沿与坐标轴平行时，则相应的投入或产出的权重为 0。

第3章 规模报酬

3.1 DEA 模型的规模报酬类型

在 DEA 模型中，对规模报酬（returns to scale, RTS）的设定决定了前沿的形状。在前一章中介绍的 CRS 和 VRS 模型，在单投入-单产出的模型中，其前沿分别是射线和凸曲线。如图所示，在 CRS 模型中，前沿为射线 OB，有效 DMU 只有 B 点一个（图 3-1）；而在 VRS 模型中，前沿为 MABD 凸曲线，有效 DMU 包括 A、B 和 D 三个（图 3-2）。

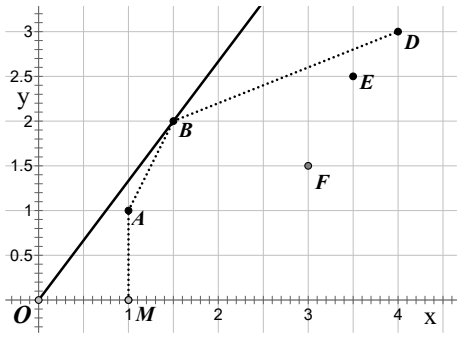


图 3-1 CRS 模型的前沿

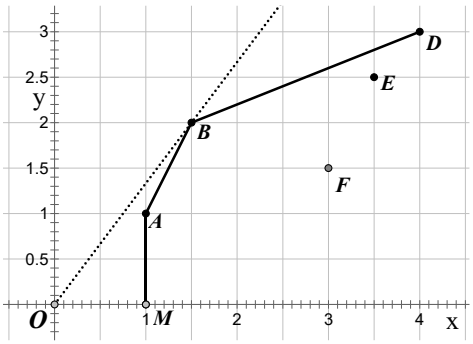


图 3-2 VRS 模型的前沿（投入导向）

那么在建立 DEA 模型时，应如何选择 RTS 类型呢？如果已知所评价的生产技术是 CRS 类型的，或者虽然所评价的生产技术是 VRS 类型的，但已知研究样本的生产规模处于 CRS 阶段，那么就应选择 CRS 模型。而在实际生产过程中，上述情况往往并不存在，生产技术的规模报酬并非 CRS，或者也不能确定研究样本的生产规模是否处于 CRS 阶段。这种情况下，如果采用 CRS 模型，得出的技术效率并非完全是纯技术效率，而是包含了规模效率成分的综合技术效率。

一般来说，生产技术的规模报酬要先后经历规模报酬递增（increasing returns to scale，IRS）、规模报酬不变(CRS)和规模报酬递减（decreasing returns to scale，DRS）三个阶段。如果无法确定研究样本处于哪个阶段，或者已知研究样本处于规模报酬的各个阶段，则评价技术效率时，应选择 VRS 模型。VRS 模型得出的技术效率是纯技术效率。

如果已知研究样本处于 IRS 或 CRS 阶段（肯定不会处于 DRS 阶段，non-decreasing returns to scale, NDRS），或者已知研究样本处于 CRS 或 DRS 阶段（肯定不会处于 IRS 阶段，non-increasing returns

to scale, NIRS）[i]，前沿分别是什么形状呢？

VRS 模型的前沿 MABD 可分为三个部分：MAB（不包括 B 点）是 IRS 阶段，B 点为 CRS 阶段，BD（不包括 B 点）为 DRS 阶段。NDRS 模型的前沿不会出现 DRS 阶段，所以 NDRS 的前沿由两部分构成：MAB（不包括 B 点）为 IRS 阶段，以 B 点为端点的射线 r 为 CRS 阶段(图 3-3)。NIRS 模型的前沿不存在 IRS 阶段，其前沿由两部分构成：OB 为 CRS 阶段，BD（不包括 B 点）为 DRS 阶段(图 3-4)。

VRS 模型是在 CRS 模型的基础上增加了约束 Σλ=1。在 VRS 前沿曲线 ABD 上的任意一点[ii]，均可在系数之和为 1 的约束下由有效 DMU（A、B、D）的线性组合构成。

在图 3-3 中曲线 MAB 及射线 r 构成了 NDRS 模型的前沿，其中在 AB 段（IRS 阶段），线性组合系数之和等于 1，在射线 r 段(CRS 阶段)，线性组合系数之和大于或等于 1（仅 B 点为等于 1），所以 NDRS 模型的线性组合系数约束为大于或等于 1，即 Σλ≥1[iii]。

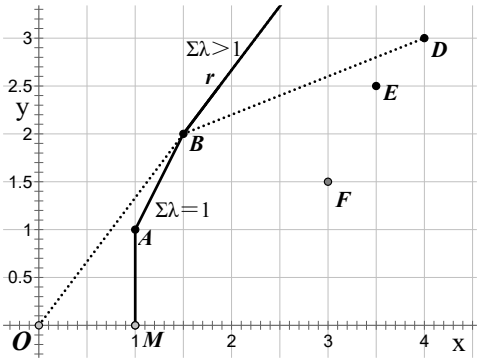


图 3-3 NDRS 模型的前沿（投入导向）

在图 3-4 中，曲线 OBD 构成了 NIRS 模型的前沿，其中在 OB 段（CRS 阶段），线性组合系数之和小于或等于 1（仅 B 点为等于 1），在 BD 段(DRS 阶段)，线性组合系数之和＝1，所以 NIRS 模型的线性组合系数约束为小于或等于 1，即 Σλ≤1。

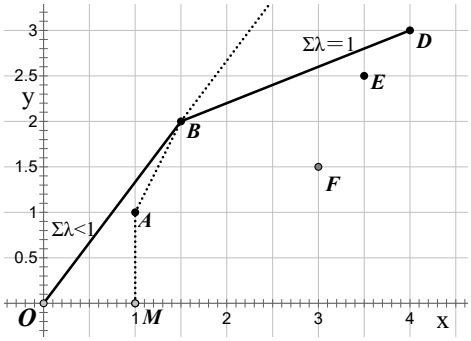


图 3-4 NIRS 模型的前沿（投入导向）

[i] 在有的文献中将 NDRS 模型称为 IRS 模型，将 NIRS 模型称为 DRS 模型，这是一种不准确的名称。因为在 NDRS 模型中，包含了 IRS 和 CRS 两个阶段，在 NIRS 模型中，包含了 CRS 和 DRS 两个阶段。

[ii] 平行于坐标轴的 MA 段为弱有效前沿，其实际的强有效投影点为 A 点，线性组合系数之和也等于 1。

[iii] 射线 OB 上的任意一点均可用 B 点乘以一个系数来构建，记做 S=kB。在 B 点以下系数 k 小于 1；在 B 点以上，系数 k 大于 1。

根据线性规划的对偶理论，可以得出各种 RTS 类型的乘数模型（表 3-1）。

表 3-1 不同 RTS 类型的 DEA 模型

	包络模型		乘数模型	
	投入导向	产出导向	投入导向	产出导向
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	$e\lambda = 1$	$e\lambda = 1$	$\mu_0\ free$	$\nu_0\ free$
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 表示一个所有元素值均为 1 的行向量，即 e=(1,1,…,1)。在不同文献的公式中，乘数模型的自由变量（μ_0 和 ν_0）前面有的用正号有的用负号，但在约束中的符号也会相应地改变，所以结果相同。

如果在研究样本中，已明确哪些 DMU 处于 IRS 阶段，哪些处于 CRS 阶段，哪些处于 DRS 阶段，则可以分组分别建立 NDRS 模型、CRS 模型和 NIRS 模型。如果只是已知生产技术为 VRS，而对研究样本中各 DMU 所处的 RTS 状态并不清楚，则选择 VRS 模型来计算纯技术效率。

本节参考文献：(Rolf Färe, Grosskopf, & Lovell, 1983; Grosskopf, 1986; William W Cooper et al., 2007; J. Zhu, 2009)。

3.2 DMU 规模报酬状态的判断

3.2.1 通过包络模型判断规模报酬状态

前一节讲的是如何根据生产技术或研究样本的规模报酬性质来选择 DEA 模型。本节讨论在已知所研究的生产技术及研究样本属于规模报酬可变时，如何判断各个 DMU 所处的规模报酬阶段，是处于 IRS、CRS、还是 DRS 阶段。

在前一节中讲到，可以采用对“线性组合系数之和”的不同约束来构建针对不同规模报酬的生产技术的 DEA 模型。当生产技术为规模报酬可变，并且研究样本所处的规模报酬阶段未知时，应建立 VRS 模型。但是在 VRS 模型的结果中，“线性组合系数之和”均为 1，显然不能通过“线性组合系数之和”的结果来判断 DMU 处于哪个规模报酬阶段。

前面讲到，在研究规模报酬可变的生产技术并且研究样本的规模报酬状态未知时，应该建立 VRS 模型。假设我们同时也建立 CRS 模型，并且把 CRS 和 VRS 模型的前沿放置于同一坐标系中，结果如图 3-5 所示，CRS 模型的前沿（射线 OC）和 VRS 模型的前沿（实线部分）放在一起。

我们观察一下，在 VRS 模型的前沿曲线上，A 点及 F 的投影点处于 IRS 阶段，B 和 C 两点及 G 的投影点处于 CRS 阶段，D 点及 E 的投影点处于 DRS 阶段。A 点和 F 点在 CRS 前沿曲线（射线 OC）上的投影点位于 B 点以下，其在 CRS 模型中的“线性组合系数之和”小于 1；B 点、C 点和 G 在 CRS 前沿曲线上的投影点位于 BC 段，其在 CRS 模型中的“线性组合系数之和”等于 1；D 点和 E 点在 CRS 前沿曲线上的投影点位于 C 点以上，其在 CRS 模型中的“线性组合系数之和”大于 1。

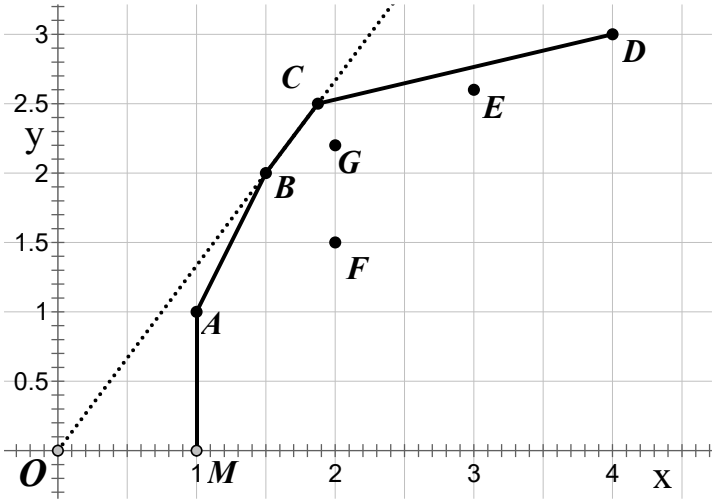


图 3-5 DMU 所处规模报酬状态的判断

总结以上规律，似乎可以很容易地借用 CRS 模型来判断 VRS 模型中各 DMU 所处的规模报酬状态：如果在被评价 DMU 的结果中，

1）$\Sigma\lambda^*<1$，则说明该 DMU 处于 IRS 状态；

2）$\Sigma\lambda^*=1$，则说明该 DMU 处于 CRS 状态；

3）$\Sigma\lambda^*>1$，则说明该 DMU 处于 DRS 状态。

但是，在 CRS 模型的前沿曲线（射线 OC）上，存在两个观测点，即 B 点和 C 点。射线 OC 上的任意一点既可由 B 点线性构造，也可由 C 点线性构造，或者由两点共同构造，即

S＝λB 或 S＝λC 或 S=λ_1B+λ_2C

这意味着在 CRS 模型中必然存在着多个最优解。在图 3-5 中，G 在 CRS 模型前沿上的投影点位于 BC 之间，如果该投影点由 B 点构造，则 λ>1；如果该投影点由 C 点构造，则 λ<1；如果由 B、C 两点共同构造，则 λ<1，λ＝1，λ>1 都可能会出现。而实际上，G 处于 CRS 状态，这说明上述判断 DMU 所处规模报酬状态的方法并不完全正确。为了避免类似的错误判断，需要将多解问题考虑在内，判断方法修正为

RTS 判断方法 1：

1）如果在 CRS 模型的所有最优解中，Σλ*<1，则说明该 DMU 处于 IRS 状态；

2）如果在 CRS 模型的所有最优解中，Σλ*>1，则说明该 DMU 处于 DRS 状态；

3）在 CRS 模型的所有最优解中，只要其中一个解 Σλ*=1，则说明该 DMU 处于 CRS 状态。

求解 DEA 模型的所有最优解并非易事。确定以上条件并不需要求解所有最优解。在实际计算中，只需要在求解 CRS 模型之后，再额外求解两个模型，分别获得 Σλ 的最大值和最小值即可。

以投入导向的模型为例，求解 Σλ 最大值（最小值）的线性规划式为

$$\max (\min) \sum_{j=1}^{n}\lambda_j$$

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

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

$$\lambda \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n \qquad (3.1)$$

在上述规划式中，如果 Max(Σλ) <1，则说明在所有最优解中 Σλ 均小于 1；如果 Min(Σλ) >1，则说明在所有最优解中 Σλ 均大于 1。

RTS 判断方法 2：

1）如果 Max (Σλ) <1，则说明该 DMU 处于 IRS 状态；

2）如果 Min (Σλ) >1，则说明该 DMU 处于 DRS 状态；

3）不在以上两种情况之列，说明该 DMU 处于 CRS 状态。

与设法获得所有最优解相比，上述求解 Σλ 最大和最小值的方法显然更为方便。但是，这种方法仍然需要额外求解两个线性规划式，增加了规划求解计算量。

根据一个最优解中的 Σλ 来判断规模报酬状态，有可能把实际处于 CRS 状态的 DMU 误判为 IRS 或 DRS。通过图 3-5 可以观察到，CRS 模型的前沿与 VRS 模型的前沿会存在重合的区域，该区域就是 VRS 前沿曲线的 CRS 部分（即 BC 段）。如果一个有效 DMU 处于 CRS 状态，则该 DMU 必定

位于前沿曲线的 CRS 区域内。结合上一章中介绍的计算规模效率的方法，可以得出，如果一个 DMU 处于 CRS 状态，则该 DMU 的规模效率＝1；反之，如果一个 DMU 的规模效率＝1，则该 DMU 处于 CRS 状态。根据这一规律，可以得出

RTS 判断方法 3：

1）如果规模效率＝1，则说明该 DMU 处于 CRS 状态；

2）如果规模效率<1，并且在任一最优解中 $\Sigma\lambda^*<1$，则说明该 DMU 处于 IRS 状态；

3）如果规模效率<1，并且在任一最优解中 $\Sigma\lambda^*>1$，则说明该 DMU 处于 DRS 状态。

3.2.2 通过乘数模型判断规模报酬状态

根据线性规划的对偶理论，规模报酬状态也可以通过乘数模型来判断。

推理 1：

以投入导向模型为例，假设 DMU K 在弱有效前沿上的投影值为（$\theta_k^* x_{ik}, y_{rk}$），则 $\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}=0$ 是通过投影点（$\theta_k^* x_{ik}, y_{rk}$）的一个前沿支撑面。

1）如果该投影点处于规模报酬不变（CRS）的状态，则将该投影点的投入和产出同时等比例增加一个很小的幅度（增加的比例为一个足够小的正数 ε）[i]，发生位移之后的投影点将仍然位于通过该点的一个支撑面上，即符合 $\sum_{r=1}^{q}\mu_r(1+\varepsilon)y_{rk}-\sum_{i=1}^{m}v_i\theta_k^*(1+\varepsilon)x_{ik}-\mu_{0k}$

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

由 $\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}=0$ 可得出 $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-(1+\varepsilon)\mu_{0k}=0$，因此 $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}=0$ 成立的条件是 $\varepsilon\mu_{0k}=0$，即 $\mu_{0k}=0$。在存在多个最优解时，只要在任一最优解中符合 $\mu_{0k}^*=0$，即说明该投影点处于 VRS 前沿的 CRS 阶段，规模报酬状态为 CRS。

2）如果该投影点处于规模报酬递增（IRS）的状态，则投影点发生位移之后将位于通过该点的支撑面的右侧，即符合 $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^* x_{ik}-\mu_{0k}<0$。其对应的条件是 $\varepsilon\mu_{0k}<0$，即

[i] 微量等比例同时增加投影点的投入和产出相当于将投影点沿原点与投影点的连线方向作微量延伸。如果该投影点处于规模收益不变（CRS）状态，则延伸点仍然位于该投影点的其中一个支撑面上（通过原点的支撑面）；如果该投影点处于规模收益递增（IRS）状态，则延伸点位于该投影点的所有支撑面的右侧；如果该投影点处于规模收益递减（DRS）状态，则延伸点位于该投影点的所有支撑面的左侧。

$\mu_{0k}<0$。如果有多个支撑面通过该投影点（即乘数模型的线性规划有多个最优解），则投影点发生位移之后将位于所有支撑面的右侧，即在所有最优解中 $\mu_{0k}^{*}<0$。

3）如果该投影点处于规模报酬递减（DRS）的状态，则投影点发生位移之后将位于通过该点的支撑面的左侧，即符合 $(1+\varepsilon)\sum_{r=1}^{q}\mu_r y_{rk}-(1+\varepsilon)\sum_{i=1}^{m}v_i\theta_k^{*}x_{ik}-\mu_{0k}>0$。如果有多个支撑面通过该投影点（即乘数模型的线性规划有多个最优解），则投影点发生位移之后将位于所有支撑面的左侧，即在所有最优解中 $\mu_{0k}^{*}>0$。

在产出导向模型中，也可做出类似推理。

推理 2：

如果边际产出大于平均产出，说明处于规模报酬递增（IRS）状态；如果边际产出等于平均产出，说明处于规模报酬不变（CRS）状态；如果边际产出小于平均产出，说明处于规模报酬递减（DRS）状态。因此通过比较被评价 DMU 在弱有效前沿上投影点的边际产出（MP）和平均产出（AP），即可判断其规模报酬状态。

现在分别将 $\sum_{i=1}^{m}v_i\theta_k^{*}x_{ik}$, $\sum_{r=1}^{q}\mu_r y_{rk}$ 看作是一个虚拟的投入和一个虚拟的产出[i]，即 $x=\sum_{i=1}^{m}v_i\theta_k^{*}x_{ik}$，$y=\sum_{r=1}^{q}\mu_r y_{rk}$。

$\sum_{r=1}^{q}\mu_r y_{rk}-\sum_{i=1}^{m}v_i\theta_k^{*}x_{ik}-\mu_{0k}=0$，简化记为 $y-x-u_{0k}=0$，是通过该投影点的一个前沿支撑面（supporting hyperplane），从而可以计算该投影点的边际产出为

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

根据

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

该投入点的平均产出为

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

[i] 对虚拟投入和虚拟产出的可理解为：乘数模型中利用权重系数构造的虚拟投入和虚拟产出对应包络模型中对投入或产出等比例变化的约束。

结合规模报酬状态的判断规则，可得出投入导向乘数模型判断规模报酬的方法（表 3-2）。

表 3-2 投入导向乘数模型规模报酬状态的判断方法

	MP 与 AP 比较	μ_0
IRS	在所有解中 MP>AP	在所有解中 $\mu_0^*<0$
CRS	在任一解中 MP＝AP	在任一解中 $\mu_0^*=0$
DRS	在所有解中 MP<AP	在所有解中 $\mu_0^*>0$

通过类似推导过程可得出，在产出导向乘数模型中 MP=1，$AP=\frac{1}{1-v_{ok}}$，从而得出判断规模报酬状态的方法（表 3-3）。

表 3-3 产出导向乘数模型规模报酬状态的判断方法

	MP 与 AP 比较	v_0
IRS	在所有解中 MP>AP	在所有解中 $v_0^*<0$
CRS	在任一解中 MP＝AP	在任一解中 $v_0^*=0$
DRS	在所有解中 MP<AP	在所有解中 $v_0^*>0$

需要注意的是，采用乘数模型方法 2 时的处理方式与包络模型的方法 2 相一致。在包络模型(3.1)中规划式的右侧是被评价 DMU 的投影值，因此，在乘数模型中，需要先对无效 DMU 做投影[i]，获得其投影值，然后求解 μ_0 或 v_0 最大和最小值。以投入导向模型为例[ii]，

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

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

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

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

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

表 3-4 总结了包络模型和乘数模型规模报酬状态的各种判断方法。MaxDEA 软件判断规模报酬状态的规则采用的是方法 3。

本节参考文献：(Banker et al., 1984; Banker & Thrall, 1992)。

[i] 此处采用弱有效投影，即原始值 + 径向改进值。

[ii] θ^*为各 DMU 的效率值，乘数模型求解的效率值与包络模型求解的效率值相等，即 $\theta_k^*=\sum_{r=1}^{q}\mu_r^* y_{rk}-\mu_0^*$。

表 3-4 规模报酬状态的判断方法总表

	包络模型		乘数模型	
	投入导向	产出导向	投入导向	产出导向
判断模型	$\min\ \theta$ $s.t.\ x_0\theta - X\lambda \geq 0$ $Y\lambda \geq y_0$ $\lambda, s^-, s^+ \geq 0$	$\max\ \phi$ $s.t.\quad X\lambda \leq x_0$ $y_0\phi - Y\lambda \leq 0$ $\lambda, s^-, s^+ \geq 0$	$\max\ y_0^T\mu - \mu_0$ $s.t.\ x_0^T\nu = 1$ $-X^T\nu + (Y^T\mu - \mu_0) \leq 0$ $\mu, \nu \geq 0$ $\mu_0\ free$	$\min\ x_0^T\nu + \nu_0$ $s.t.\ y_0^T\mu = 1$ $(X^T\nu + \nu_0) - Y^T\mu \geq 0$ $\mu, \nu \geq 0$ $\nu_0\ free$
方法 1	1）在所有最优解中 $e\lambda^*<1$，则为 IRS； 2）在所有最优解中 $e\lambda^*>1$，则为 DRS； 3）在任一最优解中 $e\lambda^*=1$，则为 CRS。		1）在所有最优解中 $\mu_0^*<0$ ($\nu_0^*<0$)，则为 IRS； 2）在所有最优解中 $\mu_0^*>0$ ($\nu_0^*>0$)，则为 DRS； 3）在任一最优解中 $\mu_0^*=0$ ($\nu_0^*=0$)，则为 CRS。	
方法 2	1）$e\lambda^*_{max}<1$，则为 IRS； 2）$e\lambda^*_{min}>1$，则为 DRS； 3）除此之外，为 CRS。		1）$\mu_0^*_{,max} < 0$ ($\nu_0^*_{,max} < 0$)，则为 IRS； 2）$\mu_0^*_{,min} > 0$ ($\nu_0^*_{,min} > 0$)，则为 DRS； 3）除此之外，为 CRS。	
方法 3	1）规模效率=1，则为 CRS； 2）规模效率<1 且任一最优解 $e\lambda^*<1$，则为 IRS； 3）规模效率<1 且任一最优解 $e\lambda^*>1$，则为 DRS。		1）规模效率=1，则为 CRS； 2）规模效率<1 且任一最优解 $\mu_0^*<0$ ($\nu_0^*<0$)，则为 IRS； 3）规模效率<1 且任一最优解 $\mu_0^*>0$ ($\nu_0^*>0$)，则为 DRS。	

X 为 m×n 投入矩阵，Y 为 q×n 产出矩阵，λ 为 n×1 列向量（参考标杆系数），ν 为 m×1 列向量（投入权重），μ 为 q×1 列向量（产出权重）

3.3 MaxDEA 求解规模效率和规模报酬状态图示

“规模报酬”选择 “**效率或 TFP 指数分解 (包含规模效率 / 规模效应)**”。



3.4 规模报酬状态与模型导向的关系

DEA 模型分析结果给出的一个 DMU 所处的规模报酬状态，实质上是其前沿投影点所处的规模报酬状态。选择不同的模型导向，会获得不同的前沿投影点，因此会得出不同的规模报酬状态。如下图所示，F 点如果采用投入导向模型得出的是规模报酬递增，而采用产出导向模型得出的则是规模报酬递减。

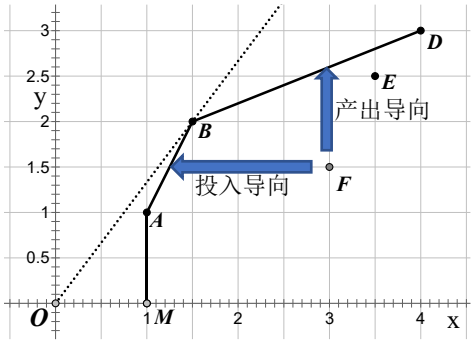


图 3-6 规模报酬状态和规模弹性与模型导向的关系

第4章DEA 模型的距离函数

在前面介绍的 DEA 模型中，无效 DMU 的改进方式为所有投入（产出）等比例缩减（增加），这类模型被称为径向 DEA 模型。径向距离函数是 DEA 模型的距离函数类型之一，除此之外，常用的距离函数类型还有至前沿最远距离函数（SBM 模型）、至强有效前沿最近距离函数、方向距离函数、至弱有效前沿最近距离函数、混合距离函数、成本函数等。

4.1 径向距离

在本书第 2 章和第 3 章已经对径向距离函数进行了详细讲解，不再重复。

在此我们补充解释两个名词：Shepherd 投入距离(Shepherd Input Distance)和 Shepherd 产出距离(Shepherd Output Distance)及其与 Farrell 技术效率的关系(Farrell, 1957)。

在投入导向径向模型(2.10)中，Shepherd 投入距离表示为 1/θ，其数值范围为 1～∞。Farrell 技术效率为 θ。

在产出导向径向模型(2.14)中，Shepherd 产出距离表示为 1/φ，其数值范围为 0～1。Farrell 技术效率也是 1/φ。

需要注意在有的文献中，分析结果报告的是投入或产出距离值，而不是 Farrell 技术效率值。在投入导向模型中，二者互为倒数；在产出导向模型中，二者相等。

当包含非期望产出时，各类导向的径向模型如下（以 VRS 为例）[i]。

1) 投入导向

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

2) 产出导向

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

3) 非导向

[i] 为了使公式更加简洁，采用矩阵形式表示。

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

$$\begin{aligned} s.t.\ & X\lambda+\beta x_k \le x_k \\ & Y\lambda-\beta y_k \ge y_k \\ & B\lambda+\beta b_k \le b_k \\ & e\lambda = 1 \\ & \lambda \ge 0 \end{aligned} \qquad (4.3)$$

投入、期望产出和非期望产出的改进比例是相等的。

4.2 至前沿最远距离（SBM 模型）

4.2.1 SBM 模型

在径向 DEA 模型中，对无效率程度的测量只包含了所有投入（产出）等比例缩减（增加）的比例。对于无效 DMU 来说，其当前状态与投影值之间的差距，除了等比例改进的部分之外，还包括松弛改进的部分。而松弛改进的部分在效率值的测量中并未得到体现。出于这样的考虑，Kaoru Tone (2001)提出了 SBM 模型（slack based measure，SBM）。

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

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

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

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

其实在此之前，J. T. Pastor, Ruiz, & Sirvent (1999)就提出了完全相同的模型，并命名为 ERM 模型（Enhanced Russell Graph Efficiency Measure）。但在后来的大多数文献中，都引用 SBM 这一名称[i]。

SBM(ERM)模型采用 ρ^*表示被评价 DMU 的效率值，它同时从投入和产出两个角度来对无效率状况进行测量，因此称为非导向（non-oriented）模型。在非导向 SBM(ERM)模型中，投入和产出数据中均不能出现 0。

在 SBM(ERM)模型中，投入和产出的无效率分别体现为

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

如果 SBM(ERM)模型的效率值（ρ^*）等于 1，则说明被评价 DMU 为强有效，不存在径向模型的弱有效问题。被评价 DMU K 的投影值（目标值）为

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

在径向模型中，无效率用所有投入（产出）可以等比例缩减（增加）的程度来测量；而在 SBM(ERM)模型中，无效率则用各项投入（产出）可以缩减（增加）的平均比例来衡量。

SBM(ERM)模型的优点是解决了径向模型对无效率的测量没有包含松弛变量的问题，但 SBM(ERM)模型也存在明显的缺点。SBM(ERM)模型的目标函数是使效率值 ρ 最小化，也就是使投入和产出的无效率值最大化。从距离函数的角度去考虑，被评价 DMU 的投影点是前沿上距离被评

[i] 此模型中的变量 s^-和 s^+均出现在目标函数中，并非松弛变量(slack)，因此将此模型命名为 SBM(slacks based measure)并不合理。

价 DMU 最远的点，这是 SBM(ERM)模型的缺点和不合理之处。从被评价者的角度，是希望以最短的路径达到前沿，SBM(ERM)模型提供的投影值显然与此相悖。

非导向的 SBM(ERM)模型是非线性规划，可按以下步骤转化为线性规划：

1）令 $\mathrm{t}=\dfrac{1}{1+\frac{1}{q}\sum_{r=1}^{q}s_r^+/y_{rk}}$，模型(4.4)转换为

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

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

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

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

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

2）令 $S^-=ts^-;S^+=ts^+;\Lambda^-=t\lambda^-$，模型进一步转换为以下线性规划

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

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

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

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

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

在模型(4.4)的目标函数中，如果分别只取分子和分母，则分别为投入导向和产出导向的 SBM(ERM)模型。

投入导向 SBM(ERM)模型为

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

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

$$Y\lambda\geq y_k$$

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

在投入导向 SBM(ERM)模型中，投入数据中不能出现 0。

产出导向 SBM(ERM)模型为

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

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

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

$$\lambda,s^+\ge 0 \qquad (4.8)$$

在产出导向 SBM(ERM)模型中，产出数据中不能出现 0。

投入导向的 SBM(ERM)模型与 Rolf Färe & Knox Lovell (1978)提出的 Russell 投入导向效率测量方法等价，但 SBM(ERM)模型的表述方式更容易理解。

表 4-1 总结了三种导向的 SBM(ERM)模型的规划式。

径向模型和 SBM(ERM)模型求解的效率值之间存在以下关系：VRS 效率值≥CRS 效率值。

在径向模型中，通过两者的数量关系，可以计算规模效率值：规模效率值＝CRS 效率值/ VRS 效率值 （SE＝TE/PTE）。或者说 CRS 效率值可以进行分解：径向 CRS 效率值＝径向 VRS 效率值* 规模效率值 （TE＝PTE* SE）。

根据规模报酬的定义，规模效率应采用径向模型计算。在 SBM(ERM)模型及其它距离函数模型中，CRS 效率值与 VRS 效率值的比值，在本书和 MaxDEA 软件中称之为“规模效应”（Scale Effect），以示区别。

对于同一被评价 DMU，其投入导向和产出导向的 SBM(ERM)效率值均大于或等于非导向 SBM(ERM)效率值。

表 4-1 各类 SBM(ERM)模型的规划式

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

4.2.2 加权 SBM(ERM)模型

在 SBM(ERM)模型中，各投入（产出）指标在效率测量中的重要程度是相同的。根据 Kaoru Tone (2001)，在 SBM(ERM)模型的基础上，可以给各项投入和产出指标赋予不同的权重，以示不同指标重要性的差异。加权处理后被评价 DMU 的有效性不受影响，即 SBM(ERM)模型和加权 SBM(ERM)模型评价为有效 DMU 是相同的，但无效 DMU 的效率值及其投影值会发生改变。

非导向加权 SBM(ERM)（weighted SBM(ERM), SBM(ERM) with preference）模型表示为

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

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

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

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

w^I 和 w^O 表示投入和产出指标的权重。

投入导向加权 SBM(ERM)模型表示为

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

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

$$Y\lambda \geq y_k$$

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

产出导向加权 SBM(ERM)模型表示为

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

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

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

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

4.2.3 MSBM 模型

MSBM(modified slack based measure)是由(Sharp, Meng, & Liu, 2007)提出的一种 SBM(ERM)模型的变种，它在目标函数中采用被评价 DMU 各项投入或产出最大可能的改进值（range of possible improvement）作为计算其无效率程度的分母，即

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

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

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

$$e\lambda=1$$

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

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

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

若 $R_{ik}^-=0$，则在目标函数中去除相应的 s_i^-/R_{ik}^- 项；若 $R_{rk}^+=0$，则在目标函数中去除相应的 s_r^+/R_{rk}^+ 项。

MSBM 与 SBM(ERM)模型相比，其优势在于允许投入和产出指标中包含负数，但其缺陷是只能应用于 VRS 模型，不适用于 CRS 模型[i]。

对投入和产出指标进行加权处理的 MSBM 模型表示为

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

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

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

$$e\lambda=1$$

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

4.2.4 MaxDEA 求解 SBM(ERM)模型和加权 SBM(ERM)模型示例

通过 MaxDEA 软件求解 SBM(ERM)模型的方法与求解径向模型的方法基本相同，不同之处包括：

1）在距离界面选择“**2) 至强有效前沿的最远距离 (ERM: Pastor, et al 1999; SBM: Tone 2001)**”。如果要运行 MSBM 模型，勾选“**调整 SBM (Sharp et al 2007)**”选项[ii]。

[i] MSBM 模型设计应用于 VRS 模型。在 CRS 和 NIRS 模型中，被评价 DMU k 投入指标的最大可能的改进值可能会超过 R_{ik}，即 $s_{ik}>R_{ik}$，从而可能使目标函数为负数。因此，MSBM 模型不适用于 CRS 和 NIRS 模型。

[ii] MSBM 仅适用于 VRS 模型，选择其他规模报酬类型可能会出现异常结果。

2）SBM(ERM)模型的乘数形式不像径向模型的乘数形式那样具有明确的含义，因此 MaxDEA 没有提供 SBM(ERM)的乘数模型，但在包络模型的结果中提供了对偶解。

我们仍然采用**表 2-6** 的数据，来比较一下径向模型和 SBM(ERM)模型的结果。

SBM(ERM)的投影点是前沿上距离被评价 DMU 最远的点，因此两种模型计算的效率值存在以下关系：

SBM(ERM)的效率值≤径向模型的效率值

在结果中，除 6 个有效 DMU 的效率值两者相等之外，SBM(ERM)模型得出的效率值均小于径向模型得出的效率值。

SBM(ERM)模型中的“距离”是指各项投入（产出）指标改进的平均比例，因此，投入（产出）导向 SBM(ERM)模型和投入（产出）导向径向模型获得的投影值中，各项投入（产出）的改进比例存在以下关系：

SBM(ERM)模型的平均改进比例 ≥ 径向模型的平均改进比例

表 4-2 径向模型与 SBM(ERM)模型的效率值和投影值比较(投入导向 CRS)[i]

	径向模型					SBM(ERM)模型				
DMU	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**
安徽	0.944	133136	125338	74967434	4391516	0.933	140997	114824	63344221	4391516
北京	0.917	80363	87385	104340626	2095434	0.798	80363	87385	104340626	2095434
福建	1.000	89947	93898	72059496	3087618	1.000	89947	93898	72059496	3087618

表 4-3 SBM(ERM)模型与加权 SBM(ERM)模型的效率值和投影值比较(非导向 CRS)

	SBM(ERM)					加权 SBM(ERM)				
DMU	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**	**效率值**	**床位数**	**卫技人员数**	**诊疗人次数**	**入院人数**
安徽	0.7038	140997	132739	116649529	4391516	0.8696	140997	116123	83884325	4391516
北京	0.7164	87596	103025	104589875	2845602	0.6075	87596	102941	104340626	2846770
福建	1.0000	89947	93898	72059496	3087618	1.0000	89947	93898	72059496	3087618

平均改进比例高并不等于每项投入（产出）的改进比例都高。以投入导向模型为例，DEA 模型的前沿为凸多面体，因此，在强有效前沿上的任意两点相比，不会存在一个点的所有投入均低于（或

[i] 原始投入产出数据见**表 2-6**。

高于）另一点的情况。在本例数据中，有两项投入，强有效前沿上的任意两点相比，如果一个点的第一种投入低于另一个点的第一种投入，则该点的第二种投入必然会高于另一个点的第二种投入。结果反映了这一规律：SBM(ERM)模型的投影点和径向模型的投影点都是前沿上的投影点（两个投影点有可能重合），其投入指标的投影值存在以下规律：如果 SBM(ERM)模型的床位数低于径向模型的床位数，则 SBM(ERM)模型的卫技人员数必然会高于径向模型的卫技人员数。

MaxDEA 软件求解加权 SBM(ERM)模型的步骤就是在 SBM(ERM)模型设置的基础上，增加权重设置。**勾选“设置投入产出权重”，并设置权重**。投入指标和产出指标的权重是相互独立的，在此将两个投入指标的权重设置为 1 和 2，两个产出指标的权重设置为 1 和 5。如果将两个投入指标的权重设置为 0.5 和 1，两个产出指标的权重设置为 2 和 10，效果是相同的。非导向 SBM(ERM)模型和加权 SBM(ERM)模型的结果比较见表 4-3。

4.3 至弱有效前沿最近距离

4.3.1 至弱有效前沿最近距离的计算方法

至弱有效前沿的最近距离（minimum distance to weak efficient frontier, MinDW）是指被评价 DMU 与前沿的最近距离，无论其在前沿的投影点是强有效还是弱有效。这种方法最早是在文献(A. Charnes, Rousseau, & Semple, 1996; Briec, 1999)中提出的，其方法可表示为 m+q 个线性规划（m 为投入指标数量，q 为产出指标数量）：

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

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

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

$$\lambda \ge 0$$

e_i 和 e_r 是常数，在规划式中只有一个 e 等于 1，其它均为 0，即

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

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

每个模型的效率值表示为 $\theta_z^*=\dfrac{1-\frac{1}{m}\sum_{i=1}^{m}\beta_z^* e_i / \mathrm{x}_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q}\beta_z^* e_r / y_{rk}}$

MinDW 模型效率值表示为 $\theta_{\max}^*=\max(\theta_Z^*,\ z=1,2,...,m+q)$，最大的效率值对应最小的 β *，即至前沿的最近距离。

如果是投入导向，则有 m 个线性规划

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

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

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

$$\lambda \ge 0$$

如果是产出导向，则有 q 个线性规划：

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

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

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

$$\lambda \ge 0$$

MinDW 模型效率值不会小于采用任何方向向量的方向距离函数模型或其它距离类型（例如径向和 SBM(ERM)）的效率值。也就是说，MinDW 模型得出的效率值是最大的。

4.3.2 MaxDEA 求解 MinDW 模型示例

MaxDEA 求解 MinDW 模型就是在距离界面选择“**3) 至弱有效前沿最近距离 (Charnes, et al 1996)**”。

采用表 2-6 中的数据，对径向模型、SBM(ERM)模型和 MinDW 模型得出的效率值的比较见表 4-4。三类模型的共同设置为产出导向，VRS。

表 4-4 不同类型距离函数模型效率值比较

地区	SBM(ERM)	径向	MinDW
安徽	0.706253	0.956817	0.977932
北京	0.817982	0.917493	0.956972
福建	1.000000	1.000000	1.000000
甘肃	0.721133	0.847287	0.916353
广东	1.000000	1.000000	1.000000

广西	0.879912	0.996415	0.998204
贵州	0.699040	0.969756	0.984646
海南	0.725478	0.932287	0.964617
河北	0.646840	0.945152	0.971803
河南	0.764488	0.969803	0.984670
安徽	0.706253	0.956817	0.977932
…	…	…	…

4.4 至强有效前沿最近距离

注意，至强有效前沿最近距离模型属于混合整数线性规划，求解耗时远超其他距离类型。

4.4.1 非导向至强有效前沿最近距离的计算方法

为方便表述，将至强有效前沿最小距离模型（minimum distance to strong efficient frontier）简称为 MinDS 模型。

既然 SBM(ERM)模型采用强有效前沿上最远的投影点存在明显的不合理之处，我们很容易想到克服这一缺陷的方法就是采用强有效前沿上距离最近的点作为投影点。

非导向至强有效前沿最近距离就是采用强有效前沿上距离最近的点作为投影点。下面的内容是讲解实现至强有效前沿最近距离的思路和计算方法。

如果在 SBM(ERM)模型(4.4)的目标函数中把最小化(min)改为最大化(max)是否就能得到强有效前沿上最近的投影点呢？

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

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

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

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

求解模型(4.14)即可知道，这一简单的想法不可行。求解这个模型的结果必然是所有松弛变量均为 0，所有被评价 DMU 的效率值均为 1，所有被评价 DMU 的参考标杆和投影值均为其自身。这是因为

$\lambda_k=1,\lambda_j=0\,(j\neq k)$，$s^-=0,s^+=0$ 是模型(4.14)的一个最优解。

如果在模型(4.14)中将参考集限定为有效 DMU，是否可行呢？即采用两阶段的方法，在第一阶段，求解 SBM(ERM)模型(4.4)，确定所有有效 DMU，记为 DMU 集合 $E=\{j\mid\rho_j=1\}$；然后在第二阶段，求解模型

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

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

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

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

我们以表 4-5 的数据为例，来求解模型(4.15)。

表 4-5 示例数据

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

第一阶段，求解 SBM(ERM)模型(4.4)，结果显示在表 4-6 中。有效 DMU 包括 A、B、C 和 D，无效 DMU 为 E、F 和 G。

表 4-6 示例数据 SBM(ERM)模型(4.4)分析结果

DMU	效率值 (ρ)	参考标杆	投影值		
			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

第二阶段，针对无效 DMU，再求解模型(4.15)，结果显示在表 4-7 中。

表 4-7 示例数据模型(4.15)分析结果

DMU	效率值 (ρ)	参考标杆	投影值		
			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

从结果中可以发现这样的问题：E 为无效 DMU，但采用模型(4.15)却得出 E 为有效 DMU 的错误结论。如果采用模型(4.4)对三个 DMU 的投影点进行检验，可以发现这三个投影点均为无效。这说明模型(4.15)也无法解决求解前沿上最近投影点的问题。产生错误的原因，可以从以下图示中看出（图 4-1）。模型(4.15)将参考集限定为 A、B、C 和 D 四个有效 DMU，从图中可以看出，即使是在约束（Σλ＝1）下，通过四个有效 DMU 的线性组合可以构建出阴影区域内的任意一点。例如 E 点可以由 A、B 和 D 的线性组合构建（在图中可理解为：先由 A 和 B 构建出 A'，然后再由 A'和 D 构建出 E），因此在模型(4.15)中原本无效的 E 被错误地判断为有效。同理，在模型(4.15)中 F 和 G 会从阴影区域内寻找距离最近的点作为投影点，最近的投影点位于 AD 连线上，从而错误地使投影点脱离前沿。

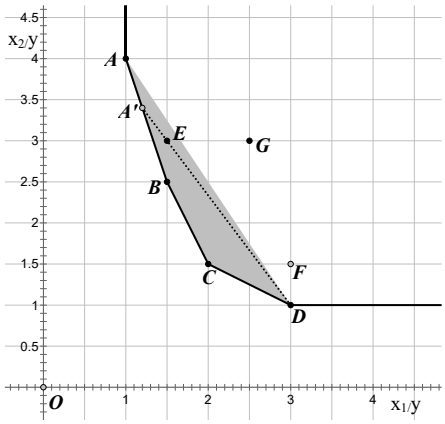


图 4-1 模型(4.15)错误示例

DEA 的前沿是由分段线性函数构成的，每一段线性函数代表一个面（facet）（在二维平面中代表一条线段），模型(4.15)产生错误的原因在于参考集中包含了来自位于不同面的有效 DMU，如果在模型中将参考集限定为同一面内的有效 DMU，则可以避免这种错误。例如在图 4-1 中，将前沿划分为 AB、BC 和 CD 三段，然后采用模型(4.15)分别依次求解无效 DMU 至三段前沿的最近距离。

遵循这样的解决思路，求解 MinDS 模型有两种方法。

方法 1：找到前沿的所有支撑面（以线性函数表示），然后分别计算无效 DMU 至各个支撑面的最小距离，最后确定其中最小的距离。找出所有支撑面的计算过程比较复杂，方法参见文献(Jahanshahloo, Lotfi, Rezai, & Balf, 2007; Jahanshahloo, Shirzadi, & Mirdehghan, 2009; F. H. Lotfi, Jahanshahloo, Mozaffari, & Gerami, 2011; Amirteimoori & Kordrostami, 2012)。SBM 模型的作者 Tone 就是按照这样的思路设计了 MinDS 模型的一种求解方法(K. Tone, 2010)。

Tone 以非导向模型为例介绍了 MinDS 模型的计算方法：

1）假设有 n 个 DMU，其中经 SBM(ERM)模型(4.4)判定为有效的 k 个 DMU 的集合记为 $E=\{j\,|\,\rho_j=1\}$ (这 k 个 DMU 是前沿所有支撑面的顶点)。

2）罗列出集合 E 的所有非空子集。用一个子集中的所有 DMU 通过线性组合构建出一个新的 DMU（线性组合系数 λ>0）。例如，如果子集 $E_1=\{A,B\}$，则新的 DMU 可以为 0.5A+0.5B。将这个新的 DMU 放入原数据中，并用 SBM(ERM)模型(4.4)检验是否为有效 DMU：如果这个新的 DMU 有效，则称该子集为集合 E 的一个**有效子集**。有效子集内的 DMU 位于前沿的同一支撑面上。重复上述过程找出集合 E 的所有有效子集。

3）用集合 E 的一个有效子集 E_z 作为参考集，求解模型(4.16)，获得效率值 ρ_z。模型(4.16)与模型(4.15)的区别在于，在模型(4.15)的参考集是由所有有效 DMU 构成的集合 E，而在模型(4.16)的参考集为集合 E 的一个有效子集。

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

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

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

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

重复这一过程，遍历所有有效子集，MinDS 模型的效率值为

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

当有效 DMU 数量很多时，通过这种方法求解最小距离的计算过程会非常繁复。根据 Tone 的推理，在 CRS 模型中，不需测试元素数量大于 m+q-1 的子集（m 为投入指标数量，q 为产出指标数量）；在 VRS 模型中，不需测试元素数量大于 m+q 的子集。如果一个子集有效，则这个子集的所有子集肯定有效，例如如果{A,B,C,D} 有效，则{A,B,C}肯定有效；如果一个子集无效，则需继续测试这个子集的所有子集，例如如果{A,B,C,D} 无效，则需继续测试{B,C,D}、{A,C,D}、{A,B,D}和{A,B,C}是否有效[i]。当有效 DMU 数量较多时，需要测试的子集数量会非常多。以 CRS 模型为例，当投入数量为 6，产出数量为 5 时，如果有效 DMU 数量为 20 个，则至少需要测试的子集数量为$C(20,10)=184756$[ii]；如果有效 DMU 数量为 30 个，则至少需要测试的全部子集数量为$C(30,10)=30045015$。由于大部分子集为无效子集，因此实际测试数量会大得多。因此，这种方法仅适用于有效 DMU 数量较少的情况。

方法 2：通过在模型(4.15)中增加约束条件，使其参考标杆位于同一超平面内。在 Tone 提出第一种求解 MinDS 模型的方法之前，J. Aparicio, Ruiz, & Sirvent (2007)就设计了更为简便的方法。该方法不需要确定所有前沿的超平面，而是通过增加约束条件，将被评价 DMU 的参考标杆限制在同一超平面内。在通过 SBM(ERM)模型确定所有有效 DMU 之后，无论有

[i] 在这些子集中，如果某个子集是有效子集的子集，则无需再测试。

[ii] 计算排列组合数量可使用 Excel 的 COMBIN 函数，$C(20,10)$ = COMBIN(20,10)。

效 DMU 数量多少，Aparicio 提出的方法都只需一个规划模型，即可求解 MinDS 模型。

Aparicio 的方法与 Tone 的方法的相同之处都是先通过 SBM(ERM)模型（或者其它等价模型）获得有效 DMU 的集合 E；不同之处在于 Aparicio 等借助于混合整数线性规划的方法，避免了对 E 的所有子集的复杂测试过程，只需要在模型(4.15)的基础上，增加一组混合整数线性约束，从而使规划模型自动以有效子集作为其参考集，巧妙地避开了测试所有子集的过程。Aparicio 的方法求解 MinDS 模型的步骤可总结为：

1）假设有 n 个 DMU，其中经 SBM(ERM)模型判定为有效的 DMU 的集合为 E；

2）求解以下混合整数线性规划，获得 MinDS 效率值

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

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

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

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

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

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

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

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

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

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

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

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

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

其中 M 是一个足够大的正数。

MinDS 模型(4.17)由三部分构成：第一部分是模型(4.15)，即目标函数和约束 a；第二部分是约束 b；第三部分是约束 c。

第二部分（约束 b）和第三部分（约束 c）的共同目的是使参考标杆位于同一个超平面内。虽然模型(4.17)并不求解投影点的超平面，也就是说，这个超平面是未知的，但是却可以通过增加约束，使参考标杆位于这个未知的超平面内。

通过对偶模型（乘数模型）可以求解出被评价 DMU 的投影点及其所有参考标杆所在的超平面。遵循这样的思路，可以通过增加乘数模型的约束，在使参考集为有效 DMU 的基础上，进一步限制其参考标杆位于同一超平面内，从而避免出现模型(4.15)中参考标杆不在同一超平面内的问题。

由于加性模型比 SBM(ERM)模型简单，不需要进行线性转换，所以采用加性模型的乘数形式（约束 b）。

虽然模型(4.15)的最优解会脱离前沿，但是在模型(4.15)所有可行解中，必然有一个解所对应的投影点是强有效前沿上距离被评价 DMU 最近的点。我们将这个解记为S^{k}_{MinDSF}，其参考标杆的集合为R^{k}_{MinDSF}。由于（松弛变量）加性模型（additive model）或加权加性模型（weighted additive model）与 SBM(ERM)模型在确定有效 DMU 集合上是等价的，因此S^{k}_{MinDSF}所对应的投影点必然也是加性模型前沿上的一个点，即S^{k}_{MinDSF}也是加性模型的一个可行解。

加性模型可表示为

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

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

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

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

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

其对偶模型（乘数形式）为

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

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

$$v_i \ge 1$$

$$\mu_r \ge 1$$

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

模型(4.15)中最小距离的可行解 S^k_{MinDSF} 也是加性模型(4.18)的一个可行解。假设该可行解在对偶模型(4.19)中对应的对偶解为 μ*和 ν*，则该可行解的所有参考标杆（R^k_{MinDSF}）所在的超平面可表示为[i]

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

下面用一个简单的数据表 4-8 来直观地演示这一规律。

表 4-8 超平面示例数据

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

在图 4-2 中，G 在前沿上的最小距离投影点为 G'，其参考标杆为 A 和 B，通过 A 和 B 的超平面为 $-5x_1 - x_2 + 10y = 0$（图中虚线表示的是 y=1 时超平面为直线 d）。除 A 和 B 之外，

[i] 通过投影点的超平面可能有多个。

其它 DMU 位于超平面的右侧，即 $-5x_1-x_2+10y<0$[i]。

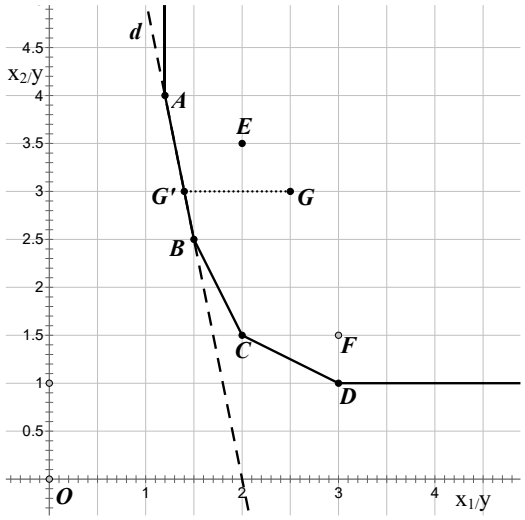


图 4-2 超平面示意图

通过以上分析，可以得出，在模型(4.15)中：如果 DMU j 位于该超平面内，则应符合 $-\sum_{i=1}^{m} v_i x_i+\sum_{r=1}^{q} \mu_r y_r=0, j\in E$。

模型(4.17)增加的两组约束就是为了满足这一规则：

1）如果 $b_j=0$，则 $d_j=0$，约束（b-1）等同于 $-\sum_{i=1}^{m} v_i x_i+\sum_{r=1}^{q} \mu_r y_r=0, j\in E$；约束（c-2）等同于 $\lambda_j <= M$，只要 M 足够大就相当于对 λ_j 的取值上限没有限制。此时 DMU_j 是 DMUk 的参考标杆。

2）如果 $b_j=1$，则 $d_j<=M$，只要 M 足够大就相当于对 d_j 取值上限没有限制，约束（b-1）等同于 $-\sum_{i=1}^{m} v_i x_i+\sum_{r=1}^{q} \mu_r y_r\le 0, j\in E$；约束（c-2）等同于 $\lambda_j=0$。此时 DMU_j 不是 DMUk 的参考标杆。

[i] 除了 A 和 B 之外，如果还有其它 DMU 位于这个超平面内，例如如果在 A 和 B 之间的连线上还有 H，则模型将存在多个最优解。投影点 G'可能由 {A，B}、{A，H}、{B，H} 或 {A，B，H} 的线性组合构建。H 点也符合 $-5x_1-x_2+10y=0$，因此在模型(4.17)中对非参考标杆的约束是 $-\sum_{i=1}^{m} v_i x_{ij}+\sum_{r=1}^{q} \mu_r y_{rj}\le 0$，而不是 $-\sum_{i=1}^{m} v_i x_{ij}+\sum_{r=1}^{q} \mu_r y_{rj}<0$。

采用模型(4.17)计算表 4-8 中 **E** 的 MinDS 模型的线性规划式如下（大 M 值＝1000000），求解结果见表 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;

表 4-9 E 点 MinDS 模型的计算结果

变量	结果
Objective（效率值）	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

在 CRS MinDS 模型(4.17)的基础上增加约束 $\sum_{j\in E}\lambda_j=1$ 和自由变量 μ_0，即为 VRS MinDS 模型：

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

需要注意的是受到线性规划求解精度的影响，M 需要是一个合适的足够大的正数，否则求解的结果可能并非正确的最优值。M 过大或过小都可能获得错误的数值。规划软件不同或者数据不同，合适的 M 值范围存在差异。

虽然 Tone 和 Aparicio 的方法都能求解最小距离函数，两者的本质思想是一致的，但显然 Aparicio 的方法更简单实用。在 MaxDEA 软件中采用的是 Aparicio 的方法。

4.4.2 MaxDEA 求解非导向 MinDS 模型示例

MinDS 距离属于混合整数线性规划，当 DMU 数量较多时，求解可能会耗时很长，软件会暂时失去响应，需耐心等待。

本节利用实际数据来说明 MaxDEA 软件求解 MinDS 模型的过程与结果。表 4-10 是 2004 年各省（市、自治区）医院的部分投入和产出指标。以床位数和卫生技术人员数作为投入指标，以诊疗人次数和入院人数作为产出指标，建立 DEA 模型。

MaxDEA 软件求解非导向 MinDS 模型采用的是方法 2，即 J. Aparicio et al. (2007)的方法。

表 4-10 各地区医院部分投入产出指标（2004 年）

地区	床位数	卫技人员数	诊疗人次数	入院人数
安徽	77773	84672	36106166	1632009
北京	69850	90656	56191348	987106
福建	54975	51611	37915799	1290919
甘肃	45123	42130	19116626	708323
广东	144416	157375	169171460	3545978
广西	64466	66521	41722576	1418468
贵州	43835	39839	17156975	820696
海南	13251	16277	7449266	249455
河北	114373	128076	48433864	2399541
河南	145870	160899	70404035	2935932
黑龙江	98975	100341	31470419	1568560
湖北	95186	107546	50361992	2036096
湖南	100167	106542	36818263	1957330
吉林	69039	78462	26756299	1082780
江苏	122787	129843	84927158	2872465
江西	58750	61535	31554864	1266984

辽宁	139635	136979	47253345	2068764
内蒙古	48074	53483	18296934	792838
宁夏	14087	14243	7323406	260350
青海	13004	12382	4781871	254352
山东	163876	181441	79416181	3785327
山西	79721	88699	28674172	1176485
陕西	77798	76225	29845077	1331931
上海	65114	70265	58799940	1141108
四川	123995	112744	65595496	2578469
天津	36645	45229	24549117	521743
西藏	4238	4918	2091942	52818
新疆	61192	60869	27318653	1316919
云南	71170	63095	34037932	1421345
浙江	102295	107777	87342404	2434372
重庆	43784	38191	23643216	823952

4.4.2.1 至强有效前沿最小距离 CRS 模型

软件操作：导入数据，设置投入产出指标，然后选择运行包络模型[i]

1）距离选择“**4) 至强有效前沿最近距离**”；

2）规模报酬选择“**1) 不变(CRS)**”。

表 4-11 是 MaxDEA 的部分输出结果。29 个无效 DMU 的 MinDS 模型效率值比 SBM(ERM)模型效率值平均高出 0.15。

表 4-11 MinDS 模型与 SBM(ERM)模型分析结果比较(非导向 CRS)

地区	SBM(ERM)		MinDS	
	效率值	参考标杆	效率值	参考标杆
安徽	0.5418	广东(0.538027)	0.7150	福建(1.414698)

[i] 软件求解 MinDS 模型所需时间会远远超过求解 SBM 模型所需的时间，MaxDEA 软件在运行过程中会暂时失去响应，特别是当数据量较大时，需耐心等待。

北京	0.5760	广东(0.483672)	0.6534	福建(0.967877); 广东(0.115229)
福建	1.0000	福建(1.000000)	1.0000	福建(1.000000)
甘肃	0.5006	广东(0.267705)	0.6420	福建(0.816299)
广东	1.0000	广东(1.000000)	1.0000	广东(1.000000)

4.4.2.2 至强有效前沿最小距离 VRS 模型

在 MaxDEA 软件包络模型的规模报酬选项中选择“**2) 可变 (VRS)**”，其它设置与 CRS 模型相同。

表 4-12 是 MinDS 的 CRS 与 VRS 模型分析结果比较。前文提到，在径向和 SBM(ERM) 模型中，CRS 模型的效率值小于或等于 VRS 模型的效率值。但是在 MinDS 模型中不存在这样的关系。**表 4-12** 中共有 15 个 DMU 的 CRS 模型效率值大于 VRS 模型效率值，这 15 个 DMU 的规模效应（CRS 模型效率值与 VRS 模型效率值的比值）大于 1。这个问题将在下一节做专门的分析。

表 4-12 MinDS 的 CRS 与 VRS 模型分析结果比较(非导向)

地区	VRS		CRS	
	效率值	参考标杆	效率值	参考标杆
安徽	0.6459	福建(0.848745); 广东(0.151255)	0.7150	福建(1.414698)
北京	0.6463	福建(0.860764); 广东(0.139236)	0.6534	福建(0.967877); 广东(0.115229)
福建	1.0000	福建(1.000000)	1.0000	福建(1.000000)
甘肃	0.6954	福建(0.437956); 青海(0.562044)	0.6420	福建(0.816299)
广东	1.0000	广东(1.000000)	1.0000	广东(1.000000)
广西	0.8612	福建(0.943439); 广东(0.056561)	0.8954	福建(1.172642)
贵州	0.7149	福建(0.546365); 青海(0.453635)	0.6740	福建(0.771909)
海南	0.8663	福建(0.127950); 青海(0.287609); 西藏(0.584440)	0.7131	福建(0.241037)

河北	0.5374	福建(0.508384); 广东(0.491616)	0.6690	福建(2.080455)
河南	0.5649	广东(0.925283); 山东(0.074717)	0.7132	福建(2.653388)
青海	1.0000	青海(1.000000)	0.6456	福建(0.236544)
山东	1.0000	山东(1.000000)	0.7574	福建(2.980919)
西藏	1.0000	西藏(1.000000)	0.5513	福建(0.077090)

4.4.3 MinDS 模型的特殊性

4.4.3.1 MinDS 模型 CRS 效率值可能会大于 VRS 效率值

径向模型和 SBM(ERM)模型及后面介绍的方向距离函数模型和 MinDW 模型求解的效率值，均存在以下关系：

VRS 效率值 ≥CRS 效率值。

在 MinDS 模型中是否还存在上述关系呢？现有文献对此没有进行讨论。

表 4-13 列出了 MinDS 模型 CRS 和 VRS 效率值及规模效应。在 31 个 DMU 中，有 15 个 DMU 的 CRS 效率值大于 VRS 效率值，依此计算的规模效应大于 1（表中黑体部分）。

表 4-13 MinDS 模型的规模效应

地区	CRS 效率值	VRS 效率值	规模效应
安徽	**0.7150**	**0.6459**	**1.1070**
北京	**0.6534**	**0.6463**	**1.0111**
福建	1.0000	1.0000	1.0000
甘肃	0.6420	0.6954	0.9232
广东	1.0000	1.0000	1.0000
广西	**0.8954**	**0.8612**	**1.0398**

这一实例说明在 MinDS 模型中，CRS 效率值会出现大于 VRS 效率值的情况。MinDS 模型的目标函数是最大化效率值（最小化距离）：1）假设 MinDS VRS 模型与 MinDS CRS 模型相比，仅仅是增加约束 Σλ＝1，则 VRS 效率值会小于 CRS 效率值（因为如果这样 VRS 模型的最优解会是 CRS 模型的一个可行解）。2）假设在 MinDS VRS 模型中删除约束 Σλ＝1，在 VRS-SBM(ERM)模型中，被评价为有效 DMU 数量一般会多于 CRS-SBM(ERM)模型，因此，在 MinDS VRS 模型参考集中的 DMU 数量会超过 MinDS CRS 模型（VRS 模型参考集包含了 CRS 模型参考集），因此 MinDS VRS 模型的效率值会大于 MinDS CRS 模型的效率值。综上所述，MinDS VRS 模型和 MinDS CRS 模型之间的差异，既有使 CRS 模型效率值更大的成分，也有使 VRS 效率值更大的成分，因此两种结果都会出现。

4.4.3.2 扩展参考集后 MinDS 模型效率值有可能增大

表 4-14 示例数据

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

假设有三个 DMU 集合，分别为 S1、S2 和 S3，例如

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

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

E_{k1}、Ek2、Ek3 表示 DMU K 分别以 S1、S2、S_3 作为参考集建立 DEA 模型计算得出的效率值。

径向模型、SBM(ERM)模型及后面介绍的方向距离函数模型和 MinDW 模型得出的效率值均存在以下关系

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

唯独 MinDS 模型不存在上述关系，当参考集扩大后，MinDS 模型得出的效率值有可能不变、减小或增大。

当扩大参考集之后，前沿的变化存在多种可能，例如：

1）前沿没有发生变化，此时各种距离函数计算得出的效率值均没有变化；

2）前沿整体往前移动，即前沿的所有部分均往前移动，此时各种距离函数得出的效率值均减小；

3）前沿没有移动，但变得更加完整，如图 4-3 所示，当参考集由 S_1 扩展到 S_3 或由 S_2 扩展到 S_3 时，前沿并没有发生移动，而是变得更加完整（由 AB 或 DE 扩展到 ABDE），在这种情况下，MinDS 模型得出的效率值有可能不变，也有可能增大。而其它距离函数模型

得出的效率值不变或减小。在 MinDS 模型中，当参考集由 S_1 扩展到 S_3 时，K 的投影点和效率值没有发生变化，投影点均为 K_1，效率值均为 0.7800；当参考集由 S_2 扩展到 S_3 时，K 的投影点由 K_2 变为 K_1，效率值由 0.7083 增大到 0.7800。

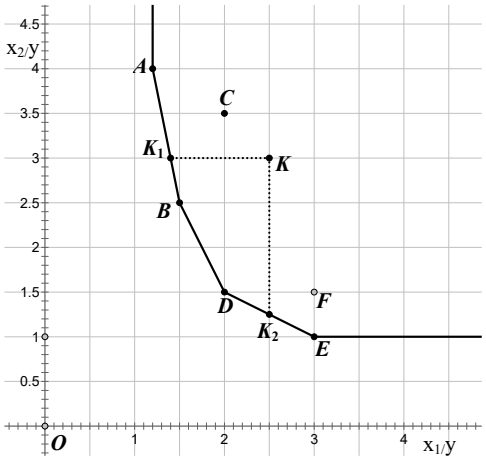


图 4-3 扩大参考集后效率值增大示意图

MinDS 模型在扩展参考集后出现效率值增大，即至强有效前沿的最小距离减小，原因在于，扩展参考集前，在被评价 DMU 的投影区域内（如图 4-4 中 K 的投影区域为 KK_1DK_2 多边形区域），强有效前沿存在缺口（K_1D 段为弱有效），当扩展参考集后，缺口被部分或完全补上，被评价 DMU 至这段补上的前沿的最小距离小于至原前沿的最小距离，从而使扩展参考集后被评价 DMU 至新前沿的距离减小。

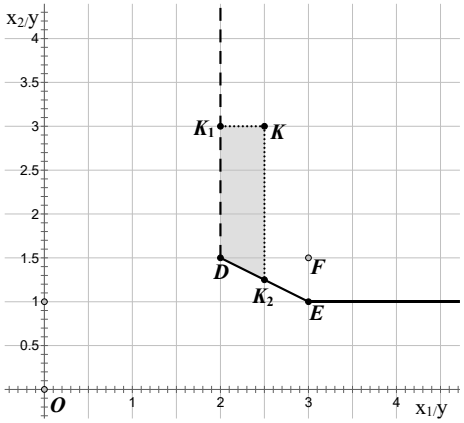


图 4-4 扩大参考集后效率值增大原因示意图

由于以上原因，MinDS 应用于以下模型时会出现异常结果：

1）序列 Malmquist 模型。序列（sequential）Malmquist 模型的一个功能是能够使得出的

技术变化指数总是大于或等于 1，也就是说，利用序列 Malmquist 模型得出的结果不会出现技术退步的情况，从而使这一模型适合于评价那些从理论上不存在技术退步的情形(Shestalova, 2003)。但是如果在序列 Malmquist 模型中采用 MinDS 距离函数，则可能会出现技术退步的异常结果。

2）共同前沿模型（MetaFrontier）。在共同前沿模型中，技术缺口比率（TGR）等于共同前沿效率值与群组前沿效率值的比值，由于共同前沿效率值小于或等于群组前沿效率值，因此得出的技术缺口比率小于或等于 1。但是如果在共同前沿模型中采用 MinDS 距离函数，则可能会出现共同前沿效率值大于群组前沿效率值的情况，从而导致技术缺口比率大于 1 的异常结果。

4.4.3.3 单调性问题

MinDS 存在的另一个缺陷是不符合单调性原则，即如果被评价 DMU 的投入减少或产出增加，得出的效率值反而可能会下降(Ando, Kai, Maeda, & Sekitani, 2012; Juan Aparicio & Pastor, 2013, 2014; Fukuyama, Maeda, Sekitani, & Shi, 2014; Ando, Minamide, Sekitani, & Shi, 2017)。

4.4.3.4 投入和产出导向 MinDS 模型

Tone 和 Aparicio 设计的求解 MinDS 模型的方法都是基于非导向的模型，尽管 Tone 在其文章最后指出，其方法能够同样应用于投入导向和产出导向的模型，但事实并非如此。无论是 Tone 还是 Aparicio 设计的求解 MinDS 模型的方法都不适用于投入和产出导向的模型。

以 CRS 模型为例，用 Tone 的方法建立的投入导向 MinDS 模型的规划式为

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

以 Aparicio 的方法建立的投入导向 MinDS 模型的规划式为

$$\max \rho_k^I = \frac{1}{m}\sum_{i=1}^{m}(1-s_i^- / x_{ik})$$

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

$$\sum_{j\in E} y_{rj}\lambda_j \geq y_k, r=1,2...,\mathrm{q}$$

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

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

$$\mu_{\mathrm{r}} \geq 0, \mathrm{r}=1,2...,\mathrm{q}$$

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

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

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

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

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

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

$$s_r^+ \geq 0, r=1,2...,\mathrm{q} \qquad (4.22)$$

以最简单的单投入-单产出的 CRS 模型为例，假设有三个 DMU，分别为 A（x=4，y=1）、B（x=3，y=3）、C（x=6，y=4），如图 4-5 所示。求解 SBM(ERM)模型(4.4)可知 B 为有效 DMU，A 和 C 为无效 DMU。然后通过模型(4.21)求解 A 的 MinDS 模型，可得其效率值＝1，这与其为无效 DMU 的结论相矛盾。图 4-5 有助于理解为什么在投入导向的模型中，通过模型(4.21)求解 MinDS 模型会得到错误的结果。在投入导向 CRS 模型中，目标函数是使 A 点与前沿的投入距离最小化，从而使得 A 在前沿的投影点为 A_I，A_I 的投入与 A 点相同，A 的投入的松弛变量 s^-=0，从而被误判为有效 DMU。前沿函数可以用从原点出发通过 B 点的射线 OB 表示，即 x-y=0。x 和 y 的取值范围均为正实数，所以所有无效 DMU 均可在射线 OB 上找到与自身投入相同的投影点，从而实现其投入松弛变量为 0。

这种投影是以牺牲产出无效率来达到使投入的无效率最小化。虽然投影点为有效 DMU，但是由于是投入导向模型，对被评价 DMU 无效率的测量忽略了产出无效率，从而使被评价

DMU 的无效率尽可能往产出上转移。例如，在图中，虽然 A_I 的投入与 A 点相同，但 A_I 的产出比 A 多，A 的产出的松弛变量 $s^+=3$，但是产出松弛变量并没有在目标函数（效率值）中体现出来。

在单投入-单产出的投入导向 CRS 模型中，无效 DMU 总是可以通过牺牲产出松弛变量的方式来获得最小的投入松弛变量值，并且最小的投入松弛变量值总是为 0。也就是说，在单投入-单产出的投入导向模型中，如果通过 Tone 或 Aparicio 的方法求解无效 DMU 的 MinDS 模型，所有无效 DMU 都会被错误地判定为有效。

同理，在单投入-单产出的产出导向 CRS 模型中，无效 DMU 总是可以通过牺牲投入松弛变量的方式来获得最小的产出松弛变量值，并且最小的产出松弛变量值总是为 0。在图 4-5 中，通过产出导向 CRS 模型，A 在前沿的投影点为 A_O，二者的产出相同，即 A 的产出松弛变量为 0。

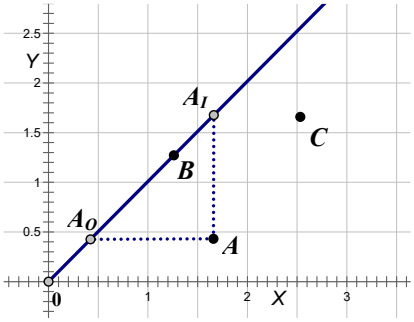


图 4-5 单投入-单产出最小距离函数示例

下面再看一个具有 2 个投入和 1 个产出的 CRS 模型的例子。如图 4-6 所示，B、C 两点为有效 DMU。根据 Tone 的证明，由 B、C 两点构成的线性组合（系数为正数）均为有效，即 $E_1=\{B,\ C\}$ 是一个有效的子集。假设以 B、C 两点作为参考集，被评价 DMU 为 G，则由式(4.21)所规定的可行域为 MBCNG，因为目标函数的方向为最大化，所以最优解为 G 自身，即规划式所确定的最优解，由 B、C 两点构成的线性组合与被评价 DMU（G）重合，效率值为 1，这显然与 G 不在前沿相矛盾。在产出导向的模型中，会出现相同的问题。

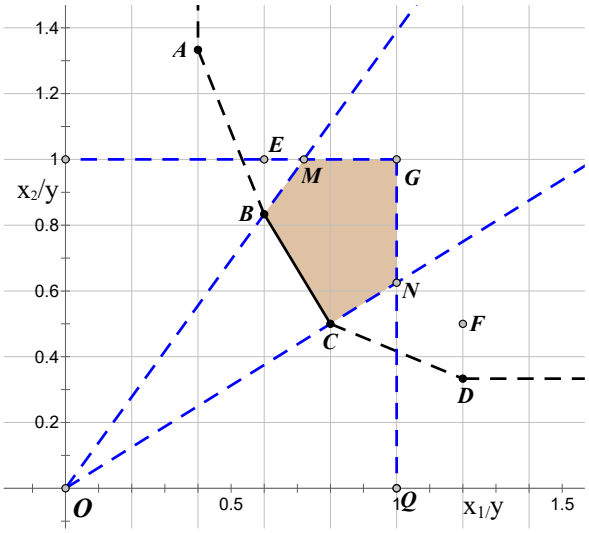


图 4-6 投入导向最大效率值求解错误示例

在 VRS 模型中，通过 Tone 或 Aparicio 的方法求解 MinDS 模型会出现同样的错误。

4.4.4 MinDS 模型算法的新进展

在前一节解释了 J. Aparicio et al. (2007)的算法只能用于求解非导向 MinDS 模型，不能求解投入导向和产出导向的 MinDS 模型。Jahanshahloo, Vakili, & Zarepisheh (2012)提出了利用双层规划求解强有效前沿最近距离的理论和方法，并且指出 J. Aparicio et al. (2007)的算法可归为双层规划方法的一个特例。虽然 Jahanshahloo et al. (2012)没有直接给出投入导向或产出导向模型的规划式，但从其给出的非导向模型很容易推导出投入导向和产出导向模型的规划式。J Aparicio, Cordero, & Pastor (2017)给出了产出导向模型的双层规划算法和等价的单层规划模型（投入导向模型的算法类似）。

针对单调性问题有两类算法。

一类是弱单调性算法。该算法是根据自由处置性原则通过扩展投影前沿来达到弱单调性(Ando et al., 2012)，例如 Q. Zhu, Wu, Ji, & Li (2018)给出的非导向模型的弱单调性算法。弱单调性是指如果被评价 DMU 的投入减少或产出增加，得出的效率值不会下降。径向距离的效率值具有弱单调性。

另一类是强单调性算法。强单调性是指如果被评价 DMU 的投入减少或产出增加，得出的效率值会提高。该算法由 Aparicio & Pastor 提出(Juan Aparicio & Pastor, 2013, 2014)，采用 Olesen & Petersen (1996)提出的 eFDEF(extended full dimensional efficient facet)作为投影前沿。实现强单调性算法有两种方法，一种是找出所有的 FDEF，然后计算无效率 DMU 与 eFDEF 的距离，例如 Juan Aparicio & Pastor (2013)的方法；另一种是采用混合整数（线性）规划(MIP)的方法，间接地实现 eFDEF 作为投影前沿，例如 Q. Zhu, Aparicio, Li, Wu, & Kou

(2022)提出的算法。

eFDEF 方法的优点是速度快，可用于计算大型数据，缺点是建模复杂；MIP 方法的优点是建模简单，缺点是速度慢，不适用于较大的数据。

表 4-15 eFDEF 方法和 MIP 方法求解产出导向强单调 MinDS 模型耗时比较

eFDEF 方法		MIP 方法	
CRS	**VRS**	**CRS**	**VRS**
4 秒	8 秒	56 分 36 秒	520 小时*

测试平台：Windows 11, CPU 为 英特尔 i9-9900K 8 核心 3.6 GHz（关闭 CPU 睿频功能，即固定运行频率为 3.6 GHz）。

计算软件：MaxDEA X 12.0，Gurobi 9.5。采用 8 线程并行计算。

测试数据：均一分布的随机数据，3 投入 3 产出，1 万 DMU。CRS 有效 DMU 数量为 61 个，VRS 有效 DMU 数量为 262 个。

*由于运行时间过长，此时间是根据求解单个 DMU 的平均耗时进行估算得出。

4.5 方向距离函数

4.5.1 方向距离函数模型

方向距离函数模型(directional distance function, DDF)是对径向 DEA 模型的一般化表达(Chung, Färe, & Grosskopf, 1997)。在方向距离函数模型中，可以由研究者自定义被评价 DMU 往前沿上投影的方向。在欧氏空间中，投影方向由方向向量决定，该方向向量由投入方向向量 v 和产出方向向量 u 构成，这是方向向量模型的最主要特征。在构成方向向量的元素中，不同投入和产出指标的方向向量值代表了其相对重要或优先程度，例如如果有两项投入，投入向量设置为（2，1），则代表在向前沿投影时，优先考虑投入 1 的改进，投入 1 每改进 2 个单位，投入 2 改进 1 个单位。

在图 4-7 投入导向方向距离函数模型示例中，v_1 和 v_2 分别定义了两个方向向量，使 G 在前沿上的投影分别为 B 点和 C 点。理论上讲，通过定义方向向量，可以使无效 DMU 沿任意设定的方向投影到前沿。

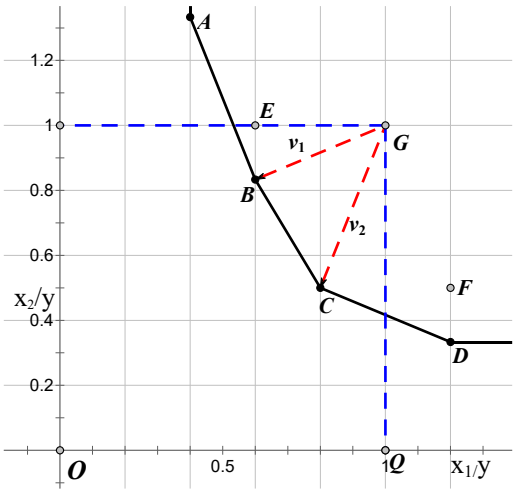


图 4-7 投入导向方向距离函数模型示例

以被评价 DMU k 为例，方向向量模型的规划式表示为

$$\max \ \beta$$

$$s.t. \ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0 \qquad (4.23)$$

在模型(4.23)中，$-g_x, g_y$ 表示投入产出方向向量，方向向量决定了被评价 DMU 往前沿上投影的方向。对无效率程度的测量（β）包含了投入和产出两个方面，因此属于非导向的方向距离函数模型。

模型(4.23)为 CRS 模型，在此基础上增加约束 Σλ＝1，即为 VRS 模型。

方向距离函数的分析结果（β）与投入和产出指标的测量单位无关，也就是说，对于同一组数据，只要方向向量不变，投入和产出指标的测量单位发生改变后，模型的结果（β）保持不变。但需要注意的一点是方向向量值是有单位的数值，其单位与对应的投入或产出指标的单位一致[i]。例如在示例数据中，各项指标的单位分别为床、人、人次、人，如果方向向量记为（1，1，1，1），是省略了测量单位，实际代表 1 个床位、1 人、1 人次、1 人。假设诊疗人次数采用“万人次”作为单位，其它指标的单位保持不变，则诊疗人次数的方向向量

[i] β 是没有单位的系数，所以方向向量值必须是有单位的，并且其单位与相应的投入或产出指标一致，否则线性规划不等式两侧将不对等。

值也要以“万人次”作为单位，变化后的方向向量为（1，1，0.0001，1），其实际代表 1 个床位、1 人、0.0001 万人次、1 人，也就是说方向向量值还是分别代表 1 个床位、1 人、1 人次、1 人。方向向量的数值如果不做上述调整，则意味着方向向量发生了改变，模型的分析结果也会相应地发生改变。

与径向模型相似，在方向距离函数模型中，也存在松弛问题。同样也可以通过两阶段方法来求解松弛变量值，即在第二阶段求解模型

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

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

$$Y\lambda - s^+ = y_k + \beta^* g_y$$

$$\lambda \geq 0$$

$$g_x \geq 0, g_y \geq 0 \quad (4.24)$$

4.5.2 方向距离函数模型的乘数形式

CRS 方向距离函数模型(4.23)的乘数形式（对偶模型）为

$$\min \sum_{i=1}^{m} v_i x_{ik} - \sum_{r=1}^{s} \mu_r y_{rk}$$

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

$$\sum_{i=1}^{m} v_i g_i + \sum_{r=1}^{s} \mu_r g_r = 1$$

$$v \geq 0; \mu \geq 0$$

$$g_i \geq 0, g_r \geq 0 \quad (4.25)$$

其它规模报酬类型的方向距离函数的乘数形式为

$$\min \sum_{i=1}^{m} v_i x_{ik} - \sum_{r=1}^{s} \mu_r y_{rk} + \mu_0$$

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

$$\sum_{i=1}^{m} v_i g_i + \sum_{r=1}^{s} \mu_r g_r = 1$$

$$v \geq 0; \mu \geq 0$$

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 方向距离函数模型中对非期望产出的处理及对弱可处置性是否合理的讨论

如果存在非期望产出（undesirable outputs，或称 bad outputs，坏产出）的情况，例如生产过程中产生的污染排放，可以在方向距离函数模型中对产出进行区分，将好产出记为 Y，坏产出记为 B，相应的产出向量区分为好产出向量 g_y 和坏产出向量 g_b。在模型中能对好产出和坏产出进行区别对待是方向距离函数模型的主要功能之一。区分好坏产出之后的模型表示为

$$\max \beta$$

$$s.t.\ X\lambda + \beta g_x \leq x_k$$

$$Y\lambda - \beta g_y \geq y_k$$

$$B\lambda + \beta g_b \leq b_k$$

$$g_x \geq 0, g_y \geq 0, g_b \geq 0 \quad (4.27)$$

$-g_x, g_y, -g_b$ 是投入、好产出和坏产出的方向向量，决定着被评价 DMU 往前沿投影的方向（减少投入、增加好产出、减少坏产出）。

模型(4.27)的生产可能集为

$$S = \{(\mathrm{x}, \mathrm{y}, \mathrm{b}) : \mathrm{x} \geq \mathrm{X}\lambda,\ y \leq Y\lambda,\ b \geq B\lambda\} \quad (4.28)$$

我们利用表 4-16 中的数据来直观地理解模型(4.27)的生产可能集，x 为投入，y 为产出，b 为非期望产出。在图 4-8 中 7 个 DMU 定义的生产可能集为 OBCD 沿纵坐标往上无限扩展的阴影区域。这个生产可能集表示，在投入既定的条件下，坏产出可以无限地增加，坏产出具有强可处置性（strong disposable）特征。

表 4-16 方向距离函数模型示例数据

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

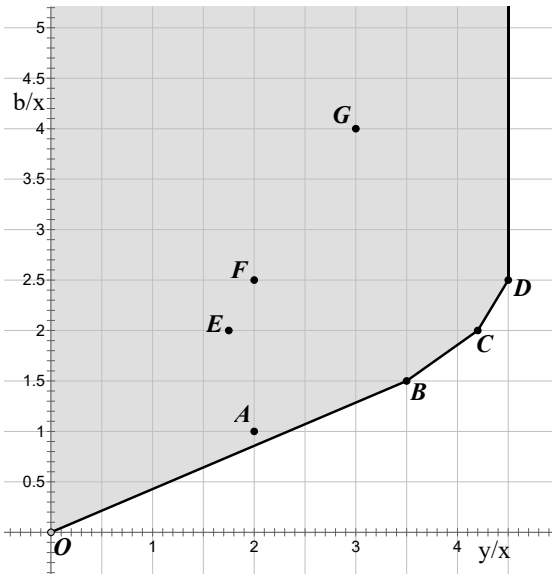


图 4-8 强可处置非期望产出方向距离函数的 PPS

非期望产出强可处置方向距离函数模型，其背后的生产可能集是不合逻辑的，在实际生产中，坏产出的数量不可能无限增加，它不应该超越生产过程所能产生的最大数量。以燃煤锅炉供暖的生产过程为例，该生产过程会排放 SO_2，SO_2 的排放数量决定于煤炭质量、生产

工艺和废气处理等因素，技术进步可以使排放数量逐步降低，这是我们所期望的结果。但反过来，从生产可能性上讲，使用质量最差的煤炭并且不进行任何的废气处理可以使 SO_2 排放数量达到最大化，但是这个最大值受到所消耗的煤炭数量（投入）的限制，在煤炭数量一定的情况下，SO_2 排放数量不可能无限增加。所以，强可处置非期望产出方向距离函数模型的生产可能集存在逻辑错误。

为了避免这一问题，Chung et al. (1997)提出的方向距离函数模型中，对生产可能集做了如下定义：

1）非期望产出具有弱可处置性（weak disposability of bad outputs），其意义为要减少坏产出，好产出也必须减少，即在投入既定的条件下，

如果 $(y,b)\in P(x)$，则 $(\theta y,\theta b)\in P(x),0\le\theta\le 1$ [i]。

2）好产出具有强可处置性（freely or strong disposability of good outputs）。即在投入既定的条件下，

如果 $(y,b)\in P(x)$，则 $(y\text{-s},b)\in P(x),s\ge 0$ 。

3）如果坏产出为 0，则好产出也为 0，即

如果 $(y,b)\in P(x)$ 且 b=0，则 y=0。

非期望产出弱可处置的方向距离函数表示为

$$\max\ \beta$$

$$s.t.\ X\lambda+\beta g_x\le x_k$$

$$Y\lambda-\beta g_y\ge y_k$$

$$B\lambda+\beta g_b=b_k$$

$$\lambda\ge 0$$

$$g_x\ge 0,g_y\ge 0,g_b\ge 0 \qquad (4.29)$$

在对非期望产出进行弱可处置性约束之后，其生产可能集缩减为图 4-9 中 OBCDGQ 所围成的有限区域（阴影区域）。

[i] $P(x)$ 表示产出集，即 $P(x)=\{(y,b):\text{x 能生产}(y,b)\}$ 。

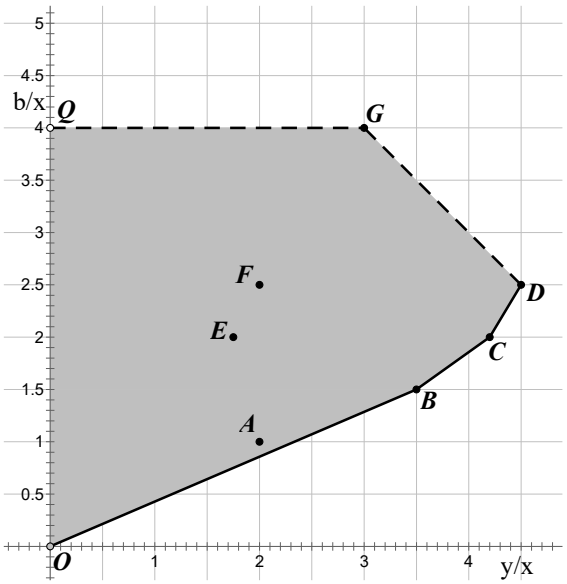


图 4-9 弱可处置非期望产出方向距离函数的 PPS

我们以 G 为例来分析弱可处置性对模型产生的影响。G 在 7 个 DMU 中，坏产出最高，但好产出低于 B、C 和 D。假设 G 拟从以下三种方案来增进其生产效率：

方案 1）保持投入和坏产出不变，通过增加好产出的途径来提高效率。要达到这样的目的，投入、好产出和坏产出对应的方向向量值可以为（0，1，0），其方向距离函数线性规划式表示为

$$\max \beta$$

$$s.t.\ X\lambda \le x_{\mathrm{G}}$$

$$Y\lambda - \beta \ge y_G$$

$$B\lambda = \mathrm{b}_G$$

$$\lambda \ge 0 \quad (4.30)$$

方案 2）保持投入不变，通过降低坏产出并同时增加好产出来提高效率，好产出增加量与坏产出减少量之比为 2：1。要达到这样的目的，投入、好产出和坏产出对应的方向向量值可以为（0，2，-1），其方向距离函数线性规划式表示为[i]

[i] 前面介绍的坏产出的弱可处置性（要减少坏产出就必须同时减少好产出）是针对生产技术的生产可能集而言的，考虑的是技术（生产前沿）的可能性。此处讲的降低坏产出并同时增加好产出是指无效 DMU 的改进方式，即无效 DMU 往前沿做投影的方向。以污染排放为例，影响生产者决策的因素不仅受到内部生产状况的影响，也受到外部因素的影响，例如法律对污染排放的处罚力度，不要将二者混淆。

$$\max \beta$$

$$s.t.\ X\lambda \le x_{\mathrm{G}}$$

$$Y\lambda - 2\beta \ge y_G$$

$$B\lambda + \beta = \mathrm{b}_G$$

$$\lambda \ge 0 \quad (4.31)$$

方案 3）保持投入不变，通过降低坏产出并同时增加好产出来提高效率，好产出增加量与坏产出减少量之比为 1：2。要达到这样的目的，投入、好产出和坏产出对应的方向向量值可以为（0，1，-2），其方向距离函数线性规划式表示为

$$\max \beta$$

$$s.t.\ X\lambda \le x_{\mathrm{G}}$$

$$Y\lambda - \beta \ge y_G$$

$$B\lambda + 2\beta = \mathrm{b}_G$$

$$\lambda \ge 0 \quad (4.32)$$

上述三个方案的分析结果见**表 4-17**。根据方案 1 和方案 2，G 的投影均为其自身，只有方案 3 的投影为前沿上 B 和 C 的线性组合。

表 4-17 弱可处置非期望产出方向距离函数的分析结果(G)

	β	参考标杆	x	y	b
方案 1	0.0000	G(1.0000)	1	3	4
方案 2	0.0000	G(1.0000)	1	3	4
方案 3	1.0526	B(0.2105); C(0.7895)	1	4.0526	1.8947

受到坏产出弱可处置性的约束，生产可能集（模型线性规划的可行域）缺少了 G 右下方的区域（图 4-10 中 GDG_2 三角形区域），因此 G 无法向这一区域做投影。

那么因坏产出的弱可处置性约束而丢失的生产可能集的 GDG_2 三角形区域，在实际生产中是否是可行的呢？

还是以燃煤锅炉供暖的生产过程为例，在同样的投入数量下，G 生产了 3 个单位的好产出和 4 个单位的坏产出，而 D 生产了 4.5 个单位的好产出和 2.5 个单位的坏产出。D 的生产

效率优于 G，主要是缘于其生产工艺的改进和对废气的处理，从生产可能性上讲，D 在保持好产出不变的情况下，通过降低对废气的处理，完全可以达到 G 的坏产出的数量水平，即如果让 D 生产 4.5 个单位的好产出，同时产生 4 个单位的坏产出是能够做到的。这说明，因坏产出的弱可处置性约束而丢失的生产可能集区域中，有部分区域实际上应该属于生产可能集。

在投入和好产出既定的条件下，在一定范围内增加坏产出，从生产技术上是可能的。

坏产出弱可处置性的约束还会造成另外一个难题，表面上看，坏产出弱可处置性 VRS 模型是在 CRS 模型的基础上增加约束 Σλ＝1，即

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda+\beta g_b=b_k$$

$$e\lambda=1$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0, g_b \ge 0 \qquad (4.33)$$

R. Färe & Grosskopf (2003)指出，模型(4.33)并非正确的坏产出弱可处置性的 VRS 模型，正确的模型应该为[i]

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$\pi Y\lambda-\beta g_y \ge y_k$$

$$\pi B\lambda+\beta g_b=b_k$$

$$e\lambda=1$$

$$0 \le \pi \le 1$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0, g_b \ge 0 \qquad (4.34)$$

但是，模型(4.34)至少存在三个方面的问题：1）它是非线性规划，比线性规划求解要困

[i] 在 MaxDEA 软件中，非期望产出弱可处置性 VRS 模型采用的是模型(4.33)，需慎重使用。

难；2）其生产可能集并非凸多面体；3）不符合单调性（monotonicity），即可能会出现被评价 DMU 增加坏产出后效率反而提高的不合理结果(Chen, 2013)。

现在我们再回头分析一下，如果在上述三种方案中，取消对坏产出的弱可处置性约束，是否会导致方向向量模型的结果产生错误呢？三种方案的线性规划模型分别为

方案 1）

$$\begin{aligned}&\max\ \beta\\&s.t.\ X\lambda\le x_{\mathrm{G}}\\&\quad Y\lambda-\beta\ge y_G\\&\quad B\lambda\le \mathrm{b}_G\\&\quad \lambda\ge 0\end{aligned}\quad (4.35)$$

方案 2）

$$\begin{aligned}&\max\ \beta\\&s.t.\ X\lambda\le x_{\mathrm{G}}\\&\quad Y\lambda-2\beta\ge y_G\\&\quad B\lambda+\beta\le \mathrm{b}_G\\&\quad \lambda\ge 0\end{aligned}\quad (4.36)$$

方案 3）

$$\begin{aligned}&\max\ \beta\\&s.t.\ X\lambda\le x_{\mathrm{G}}\\&\quad Y\lambda-\beta\ge y_G\\&\quad B\lambda+2\beta\le \mathrm{b}_G\\&\quad \lambda\ge 0\end{aligned}\quad (4.37)$$

分析结果见**表 4-18**，从结果看，G 的投影值是合乎逻辑的。方案 3 的结果与弱可处置性模型相同，方案 1 和方案 2 的投影点均为 D 点。为什么模型的分析结果并没有受到其存在逻辑错误的生产可能集的影响呢？

表 4-18 强可处置非期望产出方向距离函数的分析结果(G)

	β	参考标杆	x	y	b
方案 1	1.5000	D(1.0000)	1	4.5	2.5
方案 2	0.7500	D(1.0000)	1	4.5	2.5
方案 3	1.0526	B(0.2105); C(0.7895)	1	4.0526	1.8947

事实上，在强可处置非期望产出方向距离函数模型中，被评价 DMU 的投影方向通常是增加好产出、减少坏产出[i]。任何被评价 DMU 的坏产出的投影值都不会比其原始值更大，因此存在逻辑错误的生产可能集并不会对被评价 DMU 的投影产生影响。以 G 为例，在图 4-10 中，在追求增加好产出并减少坏产出的目标下（即好产出的方向向量为正数，坏产出的方向向量为负数），G 的投影范围为其右下的生产可能集区域，即 GG_1BCDG_2 多边形区域，由强可处置坏产出引起的生产可能集无限往上延伸的区域并不会成为被评价 DMU 的投影区域。

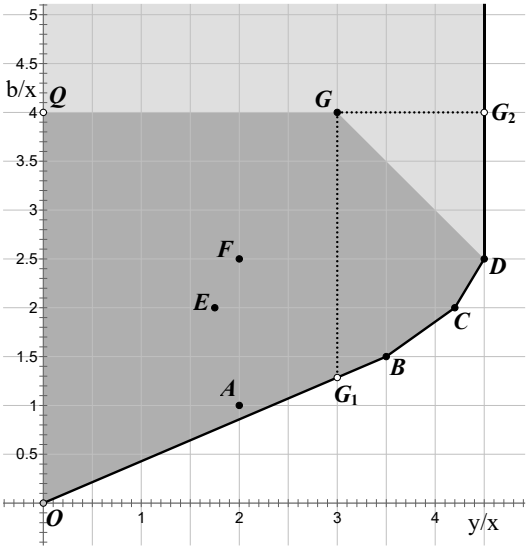


图 4-10 强可处置非期望产出方向距离函数的投影

通过以上分析，笔者认为，在方向距离函数模型中，对非期望产出进行弱可处置性约束，虽然获得广泛认可，却是一种似是而非的处理方法。一句话，虽然非期望产出强可处置模型的 PPS 是错误的，但是模型的结果是正确的。建议：在存在非期望产出时，不要对非期望

[i] 极端的情况是完全不考虑坏产出数量的多少，例如政府对污染排放没有任何处罚，在这种情况下，从生产者的角度，只会考虑增加好产出，建立的模型中不需加入坏产出，只有好产出即可，此时也就不存在坏产出是强可处置还是弱可处置的问题了。另外，理性的决策不会追求不增加好产出而只增加坏产出。如果决策者追求同时增加好产出和坏产出，即好产出和坏产出的方向向量值均为正数，在规划式中，坏产出约束的不等号方向决定了模型的最优解也通常不会超出生产技术所可能实现的范围，因为在最优解中，坏产出与好产出以一定比例同时增加，好产出不可能无限增加，所以坏产出也不会无限增加。除非，好产出和坏产出的方向向量值被设置为极不合理的数值，例如 g_y=1，g_b=非常大的不合理数值。

产出进行弱可处置性约束。

4.5.4 方向距离函数模型的几种特例

方向距离函数模型是对径向模型的推广，当方向向量取被评价 DMU 的投入产出值时，方向距离函数模型与径向 DEA 模型相同，这时方向距离函数模型中的 β^*与投入导向径向 DEA 模型效率值 θ^*的关系为 $\beta^*=1-\theta^*$，与产出导向径向 DEA 模型效率值 $1/\varphi^*$的关系为 $1/\varphi^*=1/(1+\beta^*)$。

当 $g_x=x_k$，$g_y=0$ 时，方向距离函数模型与投入导向径向模型等价：

$$\max \beta \text{ 或 } \min \theta=1-\beta$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或 } X\lambda \le (1-\beta)x_k \text{ 或 } X\lambda \le \theta x_k$$

$$Y\lambda \ge y_k$$

$$\lambda \ge 0 \quad (4.38)$$

当 $g_x=0$，$g_y=y_k$ 时，方向距离函数模型与产出导向径向模型等价：

$$\max \beta \text{ 或 } \max \phi=1+\beta$$

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

$$Y\lambda-\beta y_k \ge y_k \text{ 或 } Y\lambda_k \ge (1+\beta)y_k \text{ 或 } Y\lambda_k \ge \phi y_k$$

$$\lambda \ge 0 \quad (4.39)$$

当 $g_x=x_k$，$g_y=y_k$ 时，方向距离函数模型是非导向径向模型的一种形式：

$$\max \beta \quad \text{或} \min \frac{1-\beta}{1+\beta}$$

$$s.t.\ X\lambda+\beta x_k \le x_k \text{ 或 } X\lambda \le (1-\beta)x_k$$

$$Y\lambda-\beta y_k \ge y_k \text{ 或 } Y\lambda_k \ge (1+\beta)y_k$$

$$\lambda \ge 0 \quad (4.40)$$

当存在非期望产出时，上述三种特例如下。

当 $g_x=x_k$，$g_y=0$，$g_b=0$ 时，方向距离函数模型与包含非期望产出的投入导向径向模型等价：

$\max\ \beta$ 或 $\min\ \theta$=1-β

$s.t.\ X\lambda+\beta x_k \le x_k$ 或 $X\lambda \le (1-\beta)x_k$ 或 $X\lambda \le \theta x_k$

$Y\lambda \ge y_k$

$B\lambda \le b_k$

$\lambda \ge 0$ (4.41)

当 $g_x = 0$，$g_y = y_k$，$g_b = b_k$ 时，方向距离函数模型与包含非期望产出的产出导向径向模型等价：

$$\max\ \beta \text{ 或 } \min\ 1/\phi = \frac{1}{1+\beta}$$

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

$Y\lambda-\beta y_k \ge y_k$ 或 $Y\lambda_k \ge (1+\beta)y_k$ 或 $Y\lambda_k \ge \phi y_k$

$B\lambda+\beta b_k \le b_k$ 或 $B\lambda \le (1-\beta)b_k$

$\lambda \ge 0$ (4.42)

当 $g_x = x_k$，$g_y = y_k$，$g_b = b_k$ 时，方向距离函数模型与包含非期望产出的非导向径向模型等价：

$$\max\ \beta \text{ 或 } \min\ \frac{1-\beta}{1+\beta}$$

$s.t.\ X\lambda+\beta x_k \le x_k$ 或 $X\lambda \le (1-\beta)x_k$

$Y\lambda-\beta y_k \ge y_k$ 或 $Y\lambda_k \ge (1+\beta)y_k$

$B\lambda+\beta b_k \le b_k$ 或 $B\lambda \le (1-\beta)b_k$

$\lambda \ge 0$ (4.43)

4.5.5 一般方向距离函数模型效率值的计算方法

除上述特例以外，无法直接使用 β^*来代表无效率值。β^*受到方向向量长度的影响，例如

针对同一 DMU 的方向距离函数模型，假设方向向量采用（1,1,…）时，结果 β*=0.5，则如果方向向量采用（2,2,…），结果 β*=0.25。

$\beta \mathrm{g}_{\mathrm{x}_i}$ 表示各项投入的改进数量，$\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}$ 表示各项投入的改进比例，$\frac{1}{m}\sum_{i=1}^{m}\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}$ 表示各项投入的平均改进比例，可用来代表投入的无效率程度；$\beta \mathrm{g}_{\mathrm{y}_r}$ 表示产出的改进数量，$\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}$ 表示产出的改进比例，$\frac{1}{q}\sum_{r=1}^{q}\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}$ 表示各项产出的平均改进比例，可用来代表产出的无效率程度。因此，可以将方向距离函数效率值定义为(Cheng & Zervopoulos, 2014)：

$$\theta_k = \frac{1-\frac{1}{m}\sum_{i=1}^{m}\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q}\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}} \quad (4.44)$$

由于 β 与目标函数之间的单调变化关系，下面两种计算效率值的方式结果相同：

1）先通过求解方向距离函数模型获得最优解 β*，然后再用 β*计算效率值 θ；

2）直接采用效率值的计算公式作为方向距离函数模型的目标函数，即

$$\min \ \theta_k = \frac{1-\frac{1}{m}\sum_{i=1}^{m}\beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{q}\sum_{r=1}^{q}\beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}}$$

$$s.t.\ X\lambda + \beta g_x \le x_k$$

$$Y\lambda - \beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0 \quad (4.45)$$

投入导向方向距离函数模型的效率值为[i]

[i] 投入导向的方向距离函数与在非导向方向距离函数中，产出方向向量值＝0 等价；产出导向的方向距离函数与在非导向方向距离函数中，投入方向向量值＝0 等价。

$$\min\ \theta_k = 1 - \frac{1}{m}\sum_{i=1}^{m} \beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}$$

$$s.t.\ X\lambda + \beta g_x \le x_k$$

$$Y\lambda \ge y_k$$

$$\lambda \ge 0$$

$$g_x \ge 0 \quad (4.46)$$

产出导向方向距离函数模型的效率值为

$$\min\ \theta_k = \frac{1}{1 + \frac{1}{q}\sum_{r=1}^{q} \beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}}$$

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

$$Y\lambda - \beta g_y \ge y_k$$

$$\lambda \ge 0$$

$$g_y \ge 0 \,(4.47)$$

方向距离函数效率值的定义(4.44)具有以下特点：

1）保持了与径向 DEA 模型效率值的一致性（兼容性），当方向向量取上文中所说的与径向模型等价的三种特殊情形时，方向距离函数模型得出的效率值与径向模型相同。

2）定义(4.44)得出的方向距离函数效率值与模型中所采用的方向向量的长度无关。例如在同一模型中，方向向量取所有元素均为 1 的向量(1, 1,……, 1) 和所有元素均为 2 的向量(2, 2, ……, 2)时，得出的效率值相等。

3）在(4.44)上述定义（效率值定义）的基础上，可以方便地对投入和产出指标进行加权处理，以体现不同指标的重要程度，

$$\theta_k = \frac{1-\frac{1}{\sum w_i^I}\sum_{i=1}^{m} w_i^I \beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{\sum w_r^O}\sum_{r=1}^{q} w_i^O \beta \mathrm{g}_{\mathrm{y}_r} / y_{rk}} \qquad (4.48)$$

当存在非期望产出时，

$$\theta_k = \frac{1-\frac{1}{\sum w_i^I}\sum_{i=1}^{m} w_i^I \beta \mathrm{g}_{\mathrm{x}_i} / \mathrm{x}_{ik}}{1+\frac{1}{\sum w_r^G+\sum w_t^B}\left(\sum_{r=1}^{q} w_r^G \beta \mathrm{g}_{\mathrm{y}_r} / y_{rk} + \sum_{t=1}^{p} w_t^B \beta(-\mathrm{g}_{b_t}) / b_{tk}\right)} \qquad (4.49)$$

4.5.6 MaxDEA 求解方向距离函数模型示例

4.5.6.1 方向距离函数模型示例

以**表 2-6** 的数据为例，来说明 MaxDEA 求解一般方向距离函数模型的过程。

首先，在距离界面，选择“5) 方向距离函数 (Chambers, et al 1996; Chung, et al 1997)”。

在方向距离函数模型中，MaxDEA 提供的方向向量类型包括：

a): (-|x0|, |y0|, -|b0|)' ：这一选项就是前文介绍的方向距离函数与径向模型等价的特例，g_x, g_y, g_b 取被评价 DMU 的投入和产出数值的绝对值。

b): (-|x̄|, |ȳ|, -|b̄|)' ：g_x, g_y, g_b 取投入和产出指标平均值（均数）的绝对值。

c): (-1, 1, -1)' ：g_x, g_y, g_b 均等于 1，表示方向向量为 $\begin{bmatrix} -1 \\ \dots \\ 1 \\ \dots \\ -1 \end{bmatrix}$ 。

d): 极差 (RDM, Portela, Thanassoulis, and Simpson 2004)，是由 Portela, Thanassoulis, & Simpson (2004)提出的一种采用被评价 DMU 最大可能的改进值（range of possible improvement）作为方向向量值的方法（range directional model, RDM），即

$$g_{xk} = x_k - \min(\mathrm{x})$$

$$g_{yk} = \max(\mathrm{y}) - y_k$$

虽然 Portela et al. (2004)没有直接给出非期望产出的方向向量，但我们可以根据非期望

产出的改进方向将其方向向量值定义为 $g_{bk}=b_k-\min(b)$。

e) 自定义 (所有 DMU 使用相同向量)，自定义方向向量，定义的方向向量用于所有 DMU。点击“设置”按钮，填写各项投入和产出指标的方向向量值，填写时不需要考虑正负号，分别填写 1、1、1、15，向量值分别对应床位数、卫技人员数、诊疗人次数和住院人数。

f) 自定义 (各 DMU 使用不同向量)，为每个被评价 DMU 分别定义各自的方向向量值，即评价不同的 DMU 时，采用不同的方向向量。要采用这种方向向量，必须在导入数据前，在数据中增加各个投入和产出指标的方向向量变量（示例见表 4-19），有多少投入产出指标就相应地增加多少列向量值。这些增加的指标在数据定义时，保持“Not defined”不变。然后“设置”各个投入产出的向量变量。

在输入自定义方向向量(e 和 f)时，不需要考虑数值的正负，MaxDEA 会忽略用户输入数值的符号，并自动确定采用的正负号。

表 4-19 数据中增加方向向量指标示例

DMU	x	y	x 向量值	y 向量值
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

前面五种方向向量类型均可以通过第 6 种方向向量类型来实现，例如在导入数据前，先计算好各项投入和产出各自的最大可能的改进值，作为各自的方向向量变量加入原始数据中。然后选择第 6 种方向向量类型与直接选择第 4 种方向向量类型，即 Range (RDM: Portela et al, 2004)，是等价的。

此示例选择第 5 种自定义方向向量类型，并输入 g_x 为(1, 1), g_y 为(1, 15) [i]，表示方向向

[i] 此处方向向量值的设置仅为演示目的，没有实际意义。

量为$\begin{bmatrix}-1\\-1\\1\\15\end{bmatrix}$。

注意：

1）如果要将投入、好产出和坏产出的方向向量分离，则需要在“导向”界面选择最后一种“8) 非导向（自定义优先性）”。

2）选择投入导向与选择非导向并通过自定义将产出的方向向量值设置为 0 等价。选择投入导向时，在模型中用不到产出的方向向量值，因此产出的方向向量值不影响结果。

3）选择产出导向与选择非导向并通过自定义将投入的方向向量值设置为 0 等价。选择产出导向时，在模型中用不到投入的方向向量值，因此投入的方向向量值不影响结果。

此示例模型导向选择 Output-oriented(产出导向)；规模报酬选择 CRS。

如果要设置各项指标的权重，则在“高级模型”界面，勾选“偏好型(对投入产出进行加权)”，并设置权重[i]。

投入导向的模型，产出指标的权重没有作用；产出导向的模型，投入指标的权重没有作用。此示例模型不设置指标权重，相当于所有指标的权重相同。

分析结果见**表 4-20**。当方向向量值均大于 0 时，方向距离函数模型与径向模型在判断被评价 DMU 是否为（弱）有效状态上是等价的，但无效 DMU 的效率值及其参考标杆不同。方向距离函数模型的效率值与径向模型的效率值之间不存在固定的数量关系，在本示例模型中，大部分 DMU 的方向距离函数模型的效率值较径向模型高。

表 4-20 径向模型与方向距离函数模型分析结果比较(产出导向 CRS)

地区	方向距离函数模型		径向模型	
	效率值	参考标杆	效率值	参考标杆
安徽	0.9712	福建(0.673719); 新疆(0.825137)	0.9442	福建(0.673719); 新疆(0.825137)
北京	0.7960	福建(0.005235); 广东(0.354095)	0.9174	上海(1.000548)
福建	1.0000	福建(1.000000)	1.0000	福建(1.000000)

[i] 某些设置下，例如方向向量取第一种（被评价 DMU 自身值）时，是否设置投入产出的权重对模型不会产生任何影响（没有作用），这种情况下软件会将选项“偏好型(对投入产出进行加权)”显示为不可用状态（灰色）。

4.5.6.2 包含非期望产出的方向距离函数模型示例

下表为 2010 年各省（市、自治区）医院的床位数、卫技人员数、诊疗人次数、出院人数和出入院诊断不符人次数。将床位数和卫技人员数作为投入指标，将诊疗人次数和出院人数作为好产出指标，将出入院诊断不符人次数作为坏产出指标。坏产出指标在定义数据(Define Data)时与好产出相同，均为“Output”。

表 4-21 各地区医院部分投入产出与质量指标（2010 年）

地区	床位数	卫技人员数	诊疗人次数	出院人数	出入院诊断不符人次数
安徽	123427	123974	56090890	3717343	26021
北京	85775	121424	93376029	1708495	8542
福建	80896	82726	65632537	2710285	51495
甘肃	63773	50119	26462412	1484519	25237
广东	224114	262214	257480147	7087548	42525
广西	88913	97675	60570961	2910575	43659
贵州	69343	61260	25368195	2089306	37608
海南	18807	23324	11650197	528349	12152
河北	172956	175243	74748831	5099328	66291
河南	220974	212847	103470565	6203957	111671
黑龙江	123928	122477	45083824	2721994	24498
湖北	135006	138213	77101473	4232467	76184
湖南	150141	141921	58694543	4684786	84326
吉林	89341	83764	35533691	1994079	23929
江苏	195340	195589	150420724	5577559	33465
江西	77805	85995	42630399	2524640	37870
辽宁	160894	152822	66522624	3640376	32763
内蒙古	67016	69118	28274251	1566913	29771
宁夏	20258	20479	11314914	555371	3888
青海	16226	15519	8135941	406122	6092
山东	255764	258707	123570567	7533675	75337
山西	108260	115090	36443520	2132218	17058
陕西	104819	111967	49244614	2849597	31346
上海	84825	93566	104898763	2114263	10571
四川	184828	172381	98207483	5557053	83356
天津	40387	50897	43334154	933063	9331
西藏	5444	5435	3390468	108864	2177
新疆	89871	78206	34160068	2644559	29090
云南	112493	83687	57102579	3316600	89548
浙江	150986	171461	156681377	4345922	65189
重庆	64827	60294	34036573	1799571	23394

当存在非期望产出时，距离、导向、规模报酬和指标权重的设置方法与一般方向距离函数模型相同。

此示例模型采用第 1 种方向向量类型，即 “**a) =（-x_0, y_0, -b_0）**”；导向选择“**3) 非导向**”；规模报酬选择 VRS。除此之外，在“高级模型”界面，勾选“**非期望产出**”选项[i]。

如果通过自定义方向向量（第 5 和第 6 种方向向量类型）将一个产出指标的方向向量值设置为负数，但没有将其定义为坏产出，则软件实际计算的是错误的模型(4.27)。

如果将非期望产出设置为弱可处置（weak disposable），则在“高级模型”界面，勾选“**弱可处置**”选项，并将非期望产出设置为弱可处置。

此示例模型分别设置非期望产出为强可处置（不勾选“**弱可处置**”选项）和弱可处置，对其结果进行比较[ii]。

分析结果见表 4-22。由于坏产出弱可处置模型的最优解是强可处置模型的一个可行解，因此，弱可处置模型的效率值大于或等于强可处置模型的效率值。在本示例中，有 5 个 DMU 的弱可处置模型的效率值大于强可处置模型，其中有 3 个 DMU（河南、湖北、浙江）在强可处置模型中为无效，而在弱可处置模型中为有效。

表 4-22 包含非期望产出的方向距离函数模型分析结果

地区	强可处置模型		弱可处置模型	
	效率值	β	效率值	β
广西	0.9984	0.0008	0.9984	0.0008
贵州	0.9902	0.0049	0.9902	0.0049
海南	**0.9323**	**0.0350**	**0.9885**	**0.0058**
河北	0.9543	0.0234	0.9543	0.0234
河南	**0.9508**	**0.0252**	**1.0000**	**0.0000**

[i] 采用第 1 种方向向量类型时，即 “**a) =（-x_0, y_0, -b_0）**”，“好产出的总权重(w_g)”和“坏产出的总权重(w_b)”两个参数不发挥作用（设置与否对结果没有影响），因此软件将其显示为不可用状态。当选择其他方向向量类型时，可以对这两个参数进行设置。

[ii] 笔者建议将非期望产出设置为强可处置，原因见前文分析。

4.5.7 对包含非期望产出 DEA 模型的结果容易产生的误解

以下是常见的对包含非期望产出 DEA 模型结果的一种误解。

模型 1：不包含非期望产出的径向或方向距离函数模型；

模型 2：在模型 1 的基础上增加一个或多个非期望产出指标。

分析结果发现，模型 2 得出的效率值总是大于或等于模型 1 得出的效率值（模型 2 的效率值可能保持不变，但不会出现效率值下降的情况）。对于这样的结果，许多初学者往往难以理解。在增加了非期望产出之后，效率值不是应该下降吗？

这种误解源于对“技术效率”和“生产率”两个概念的混淆。若是对生产率进行测量，包含非期望产出时得出的生产率是低于不包含非期望产出时得出的生产率的。但技术效率并非生产率，技术效率是相对（于生产前沿的）效率，可以看作是被评价 DMU 的生产率与生产前沿的生产率之比值，其含义类似于生产率指数。生产率高，生产率指数未必高，反之亦然。当包含非期望产出时，得出的技术效率值更高，是因为被评价 DMU 的生产率的下降比例低于技术前沿的生产率的下降比例，从而使二者的比值（技术效率）升高。

通过 DEA 方法是无法计算生产率（或者全要素生产率，即 TFP）的，只能计算 TFP 指数（详见第 8 章）。通过全局或固定参比方式，可以计算累积的 TFP 指数。累积 TFP 指数仍然是指数，累积计算改变的只是相对期，即由 t 期相对于 t-1 期，变为 t 期相对于 t_0 期（基期）。累积计算和直接计算 t 期相对于 t_0 期的 TFP 指数，得出的结果是一样的。无论是 TFP 指数还是累积 TFP 指数，都只能反映 TFP 的变化，不能得出 TFP 本身数值的大小。

下面从计算过程来看为什么在增加了非期望产出之后，得出的技术效率值会提高。

前面我们曾经说过，当 DMU 数量固定时，纳入模型的投入产出指标数量越多，得出的效率值越大。在原来模型基础上（无论是 CCR 还是 BCC 模型），如果增加一项投入或产出指标，新模型得出的效率值一定是大于或等于原模型得出的效率值。再进一步说，无论增加的指标是投入还是产出，无论是期望产出还是非期望产出，新模型得出的效率值一定是大于或等于原模型得出的效率值。此结论同样适用于方向距离函数模型。这是对两个 DEA 模型之间的比较。

我们可以简单地来证明以上结论。

模型 1：

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$g_x \ge 0, g_y \ge 0 \quad (4.50)$$

模型 2：

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda+\beta g_b \le b_k$$

$$g_x \ge 0, g_y \ge 0, g_b \ge 0 \quad (4.51)$$

模型 2 与模型 1 相比，唯一的差别是模型 2 增加了一个或多个针对非期望产出的约束项 $B\lambda-\beta g_b \ge b_k$。模型的目标函数是最大化 β ，模型 2 在增加了一个或多个约束项之后所能获得的 β 最大值肯定不会超过模型 1 得出的 β 最大值。β 值越大，效率值越小，因此模型 2 得出的效率值肯定是大于或等于模型 1 得出的效率值。

认为在增加了非期望产出之后，技术效率值应该下降的想法，其真正的含义应该是针对某一个特定的 DMU 而言的。准确的表述应该是：在包含非期望产出的模型中，如果某一个 DMU 其非期望产出数量增加(假定其它 DMU 保持不变)，则该 DMU 的效率值是下降的（实际结果也可能保持不变，但肯定不会提高）。我们可以在包含非期望产出的模型中，改变某一个 DMU 的非期望产出的数值，对以上结论进行验证。

MaxDEA 提供了更为简单的方法，不需要一次一个 DMU 地进行分析。以表 4-23 中的数据为例，两项投入(X1, X2)，一项期望产出(Y)，一项非期望产出(B)。0 组是原始数据，1 组是非期望产出减少，2 组是非期望产出增加。示例采用方向距离函数非导向 VRS 模型，方向向量类型采用“**a) = (-x_0, y_0, -b_0)**”[i]。在 MaxDEA 软件“高级模型”界面，勾选“非期

[i] 等价于径向距离。

望产出”并将 B 设置为非期望产出；选择“群组参比”，“h) 固定参比”，参比群组选择 0 组[i]。

表 4-24 是分别减少和增加非期望产出后效率值变化，示例数据的结果是，减少非期望产出之后效率值增加，增加非期望产出之后，效率值减小或不变。

表 4-23 非期望产出示例数据

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

[i] “固定参比”是 8.12 版本新增加的一种群组参比方法。正在使用 8.0 - 8.11 的用户，在软件内点“检查更新”可免费更新至 8.x 最新版本。使用 7.x 或更早版本的用户，在“高级模型”界面，选择“自定义参比”，将 0 组的 DMU 设置为参考集，可以获得同样的结果。

表 4-24 分别减少和增加非期望产出后效率值变化

DMU	原始效率值	非期望产出减少后		非期望产出增加	
		1 组效率值	效率值变化	2 组效率值	效率值变化
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 加权加性距离（包括 SBM 方向距离函数和非径向方向距离函数）

本节模型均以 VRS 为例。

本节所介绍的各类模型，均属于同一距离类型，加权加性距离。它们是采用了不同权重的加性模型，其中 RAM 和 BAM 可以看作是方向 SBM(SBM 方向距离函数)和非径向方向距离函数的特例，即 RAM 和 BAM 模型可以看作是采用了特定伪方向向量值的方向 SBM 或非径向方向距离函数（下文详述）。

除 RAM 和 BAM 模型的结果为“效率值”(Score)之外，其他类型的加权加性模型结果均为“无效率值”(Inefficiency Score)，其数值范围为 0-无穷大，如果是超效率模型，则“无效率值”的数值范围为实数（即负无穷大-正无穷大）。同一数据，得出的“无效率值”的实际数值范围取决于设置的权重。权重数值越小，“无效率值”越小。“无效率值”无法用于计算 Malmquist 指数，只能计算 Luenberger 指数，详见第 8 章。如果需要计算 Malmquist 指数，选择 RAM 或 BAM 模型。

4.6.1 不加权的加性距离（简单加性距离）

不加权的加性模型的目标函数是各投入产出指标的松弛变量的直接加总，其表达式为

$$\begin{aligned}
&\max \sum_{i=1}^{m} s^{-} + \sum_{r=1}^{q} s^{+} \\
&s.t.\ X\lambda + s^{-} = x_k \\
&\quad Y\lambda - s^{+} = y_k \\
&\quad e\lambda = 1 \\
&\quad \lambda, s^{-}, s^{+} \geq 0 \qquad (4.52)
\end{aligned}$$

上面的表达式存在着逻辑错误。各投入产出的松弛变量是有单位的，不同单位的松弛变量直接相加逻辑上是不通的。

把模型(4.52)修正为下节中的简单加权的加性模型（权重数值＝1），则可以消除以上逻辑错误。

当模型包含非期望产出时，

$$\max \sum_{i=1}^{m} s^{-} + \sum_{r=1}^{q_1} s^{+} + \sum_{t=1}^{q_2} s^{b-}$$

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

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

$$\mathrm{B}\lambda + s^{b-} = b_k$$

$$e\lambda = 1$$

$$\lambda, s^{-}, s^{+} \geq 0 \qquad (4.53)$$

在此类模型中，如果将非期望产出作为投入处理，得到的结果是相同的。

4.6.2 加权加性距离

加权加性模型是在加性模型基础上，在目标函数中，对各松弛变量赋以不同的权重 w，即目标函数为各松弛变量的加权之和，表达式为

$$\max \sum_{i=1}^{m} w_i^{-} s_i^{-} + \sum_{r=1}^{q} w_r^{+} s_r^{+}$$

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

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

$$e\lambda = 1$$

$$\lambda, s^{-}, s^{+} \geq 0 \qquad (4.54)$$

注意在上面的表达式中，松弛变量的权重 w 是有量纲的，其量纲为相应投入或产出量纲（数值单位）的倒数。例如，假设 x_i 的单位为 kg，则 w_i 的单位为 1/kg。

当所有权重 w 的数值部分均等于 1 时，就简化为“简单加权加性模型”。由于权重是有量纲的，因此简单加权加性模型不存在上面提到的不加权加性模型的逻辑错误。

单位不变性（unit invariance）是指当改变投入或产出的数据测量单位后，模型的结果保持不变。未经修正的加性模型，即模型(4.52)(编号)，不符合单位不变性，假设某项投入原来的单位是千克 (kg)，当改为克(g)时，该项投入的数值将扩大为原来的 1000 倍，数据单位改变后，模型(4.52)的分析结果会随之改变。

加权加性模型由于权重包含了量纲，当数据单位改变后，模型结果将保持不变。假设某

项投入原来的单位是千克 (kg)，权重为 0.5/kg，当改为克(g)时，其权重 0.5/kg＝0.5/(1000g) ＝ 0.0005/g，模型(4.54)的分析结果将保持不变。当数据单位改变时，其权重的数值部分和量纲部分同时发生相应的改变，数值部分由 0.5 变为 0.0005，量纲部分由 1/kg 变为 1/g，实际权重保持不变。

注意在软件中设置的权重，只是权重的数值部分，如果数据的测量单位发生变化，用户应对权重的数值部分进行相应的修改。如果不修改权重的数值部分，则意味着实际权重将发生变化，分析结果会随之改变。

如果在加权加性模型中包含非期望产出，则表达式为

$$\max \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q_1} w_r^+ s_r^+ + \sum_{t=1}^{q_2} w_t^{b-} s_t^{b-}$$

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

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

$$\mathrm{B}\lambda + s^{b-} = b_k$$

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0 \qquad (4.55)$$

可以看出，在包含非期望产出的加权加性模型中，如果将非期望产出作为投入处理，得到的结果是相同的。

在 MaxDEA 软件中，用户可以自定义权重 w。

h) 自定义 (所有 DMU 使用相同权重)

i) 自定义 (各 DMU 使用不同权重)

h 是直接填写权重，i 是在原始数据中增加相应的权重变量。

使用 h 分项时，直接为每项投入产出填写权重数值（软件默认的权重数值均为 1），设置的权重适用于所有 DMU；

使用 i 分项时，为每项投入产出选择权重变量，因此可以为不同 DMU 设置不同的权重。

使用 i 分项时，必须在导入数据之前，在数据中为每项投入和产出增加相应的权重变量。注意权重变量在定义数据时，保持“Not defined”不变。

除了自定义权重之外，还可以在几种常见的权重类型中选择使用一种，包括：

b）标准化权重(Lovell & Pastor, 1995)

采用变量的样本标准差（Sample Standard Deviation）的倒数作为权重

c）权重 = $1/|x_0|, 1/|y_0|$

采用被评价 DMU 自身数值绝对值的倒数作为权重

d）权重 = $1/|\bar{x}|, \quad 1/|\bar{y}|$

采用变量均数绝对值的倒数作为权重

e）极差调整（RAM）(W. W. Cooper, Park, & Pastor, 1999)

f）边界值调整(BAM)(William W. Cooper, Pastor, Borras, Aparicio, & Pastor, 2011)

g）方向 SBM (Fukuyama & Weber, 2009)

RAM、BAM 和方向 SBM 在后面单独介绍。

4.6.3 加权加性距离之简单加权的变形表达：基于方向距离函数之和的 SBM（slacks-based measure on the sum of directional distance functions）

基于方向距离函数之和的 SBM 是由 R. Färe & Grosskopf (2010)提出的一种简单加性模型的变形表达方式，实质上为权重数值为 1 的加权加性模型。该模型表示为

$$\max \ \beta_1^x + \ldots + \beta_m^x + \beta_1^y + \ldots + \beta_q^y$$

$$s.t. \ \sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x e_i^x \le x_{i0}, i = 1,2,3,\ldots,m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y e_r^y \ge y_{r0}, r = 1,2,3\ldots,q$$

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

$$\lambda, \beta^x, \beta^y \ge 0 \qquad (4.56)$$

e 由两部分组成，包括数值部分和量纲部分，其数值部分均为 1，量纲与对应的投入或

产出相同。β 无量纲。

注意模型(4.56)中不等号改为等号后，模型结果保持不变[i]，即

$$\max \ \beta_1^x + \ldots + \beta_m^x + \beta_1^y + \ldots + \beta_q^y$$

$$s.t. \ \sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x e_i^x = x_{ik}, i = 1,2,3,\ldots,m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y e_r^y = y_{rk}, r = 1,2,3\ldots,q$$

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

$$\lambda, \beta^x, \beta^y \geq 0 \qquad (4.57)$$

令 s = βe，w =1/e，则模型(4.57)可转化为

$$\max \ \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q} w_r^+ s_r^+$$

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

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

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \geq 0 \qquad (4.58)$$

转换后的模型与加权加性模型(4.54)完全相同。

因此在 MaxDEA 软件中，要实现基于方向距离函数之和的 SBM 模型，只需选择加权加性

---

[i] 以其中第 m 项投入的约束为例来说明在此模型中不等号与等号等价。

$\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x \leq x_{m0}$ 等价于 $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x + \sigma_m = x_{m0}$，$\sigma_m \geq 0$。由于目标函数是最大化各项 β 之和，假设在模型一个解中存在 $\sigma_m^* > 0$，则一定可以在保持 $\beta_i \ (i \neq m)$ 的最优解 $\beta_i^* \ (i \neq m)$ 不变的条件下，得到一个更优解 $\beta_m^{*2} = \beta_m^* + \sigma_m^*, \sigma_m^{*2} = 0$。因此，在模型的最优解中，必然存在 σ_m^*=0。从而得出结论 $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x \leq x_{m0}$ 等价于 $\sum_{j=1}^{n} x_{mj}\lambda_j + \beta_m^x e_m^x = x_{m0}$。

模型下第 1 个分选项，即“ a) 简单加权： 权重 = (1, 1, 1, …)”即可。

4.6.4 加权加性距离之 RAM 模型

RAM（Range Adjusted Measure）模型是由 W. W. Cooper et al. (1999)提出的一种加权加性模型，与其他加权加性模型的一个显著区别是，RAM 模型定义的是效率值，且数值介于 0-1 之间[i]。

$$\min\ 1-\frac{1}{m+q}\left(\sum_{i=1}^{m}\frac{s_i^-}{R_i^-}+\sum_{r=1}^{q}\frac{s_r^+}{R_r^+}\right)$$

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

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

$$e\lambda=1$$

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

$$R_i^-=\max(\mathrm{x}_i)-\min(\mathrm{x}_i)$$

$$R_r^+=\max(\mathrm{y}_r)-\min(\mathrm{y}_r) \qquad (4.59)$$

当模型包含非期望产出时，

[i] RAM 模型设计应用于 VRS 模型。在 CRS 和 NIRS 模型中，被评价 DMU k 投入指标的最大可能的改进值可能会超过 R_i，即 $s_{ik}>R_i$，从而可能使目标函数为负数。因此，RAM 模型不适用于 CRS 和 NIRS 模型。

$$\min\ 1-\frac{1}{m+q_1+q_2}\left(\sum_{i=1}^{m}\frac{s_i^-}{R_i^-}+\sum_{r=1}^{q_1}\frac{s_r^+}{R_r^+}+\sum_{t=1}^{q_2}\frac{s_t^{b-}}{R_t^{b-}}\right)$$

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

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

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

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

$$R_i^-=\max(\mathrm{x}_i)-\min(\mathrm{x}_i)$$

$$R_r^+=\max(\mathrm{y}_r)-\min(\mathrm{y}_r)$$

$$R_t^{b-}=\max(b_t)-\min(b_t) \qquad (4.60)$$

在 RAM 模型中，如果将非期望产出作为投入处理，得到的结果是相同的。

4.6.5 加权加性距离之 BAM 模型

BAM（Bounded Adjusted Measure）模型是由 William W. Cooper et al. (2011)提出的一种与 RAM 模型类似的加权加性模型，其效率值也是介于 0-1 之间。[i]

[i] BAM 模型设计应用于 VRS 模型。在 CRS 和 NIRS 模型中，被评价 DMU k 投入指标的最大可能的改进值可能会超过 L_{ik}，即 $s_{ik}>L_{ik}$，从而可能使目标函数为负数。因此，RAM 模型不适用于 CRS 和 NIRS 模型。

$$\min\ 1-\frac{1}{m+q}\left(\sum_{i=1}^{m}\frac{s_i^-}{L_{ik}^-}+\sum_{r=1}^{q}\frac{s_r^+}{U_{rk}^+}\right)$$

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

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

$$e\lambda=1$$

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

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

$$U_{rk}^+=\max(\mathrm{y}_r)-y_{rk} \qquad (4.61)$$

若 $L_{ik}^-=0$，则在目标函数中去除相应的 s_i^-/L_{ik}^- 项；若 $U_{rk}^+=0$，则在目标函数中去除相应的 s_r^+/U_{rk}^+ 项。

$L_{ik}^-\leq R_i^-$，$U_{rk}^+\leq R_r^+$，因此 BAM 模型效率值≤RAM 模型效率值。

当模型包含非期望产出时，

$$\min\ 1-\frac{1}{m+q_1+q_2}\left(\sum_{i=1}^{m}\frac{s_i^-}{L_{ik}^-}+\sum_{r=1}^{q_1}\frac{s_r^+}{U_{rk}^+}+\sum_{t=1}^{q_2}\frac{s_t^{b-}}{L_{tk}^{b-}}\right)$$

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

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

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

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

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

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

$$L_{tk}^{b-}=b_{tk}-\min(\mathrm{b}_t) \qquad (4.62)$$

在 BAM 模型中，如果将非期望产出作为投入处理，得到的结果是相同的。

4.6.6 加权加性距离之方向 SBM (directional slacks-based measure)

方向 SBM（也被称为 SBM 方向距离函数）是 Fukuyama & Weber (2009)提出的一种模型，该模型的表达式为

$$\begin{aligned}\max\ & \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+ / g_r^y)}{2}\\ s.t.\ & X\lambda+s^-=x_k\\ & Y\lambda-s^+=y_k\\ & e\lambda=1\\ & \lambda,s^-,s^+\geq 0\end{aligned} \qquad (4.63)$$

（g^x, g^y）是伪方向向量。注意 g^x, g^y 都是正数，且是包含量纲的，其量纲与对应的投入或产出相同，因此方向 SBM 也符合上文讨论的单位不变性。

如果将其目标函数改变一下形式，就会发现方向 SBM 本质上是一种加权加性模型。

$$\begin{aligned}&\max \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+ / g_r^y)}{2}\\ &=\frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^- / g_i^x)+\frac{1}{q}\sum_{r=1}^{q}(s_r^+ / g_r^y)}{2}\\ &=\sum_{i=1}^{m}\frac{1}{2mg_i^x}s_i^-+\sum_{r=1}^{q}\frac{1}{2qg_r^y}s_r^+\end{aligned} \qquad (4.64)$$

可以看出，方向 SBM 模型就是权重为 $w^x=\frac{1}{2mg^x}$ 和 $w^y=\frac{1}{2qg^y}$ 的加权加性模型。

MaxDEA 在加权加性模型下直接提供了方向 SBM 模型分选项，即“6) 加权加性距离(Weighted Additive)”下面的分项 “g) 方向 SBM (Fukuyama & Weber, 2009)”。通过此分项，用户可以更方便地构建方向 SBM 模型，只需直接设置伪方向向量值（g^x 和 g^y）即可。MaxDEA 提供了多种伪方向向量类型，可以直接选择，用户也可以自定义伪方向向量。

其中第 4 项“(d) 极差 (RDM, Portela, Thanassoulis, and Simpson 2004)”是借鉴 Portela et al. (2004)所提出的 RDM 方向距离函数方法，在这里伪方向向量为被评价 DMU

最大可能的改进值（range of possible improvement），即

$$g_{xk}=x_k-\min(\mathrm{x})$$

$$g_{yk}=\max(\mathrm{y})-y_k$$

针对非期望产出，$g_{bk}=b_k-\min(b)$

如果规模报酬选择 VRS，则这种伪方向向量类型得出的无效率值将介于 0-1 之间，可以比较方便地转换为效率值，例如采用 “效率值 = 1 - 无效率值”。

如果采用自定义伪方向向量（e 和 f），注意无论是投入还是产出，无论是期望产出还是非期望产出，其伪方向向量值都是正数。

当模型包含非期望产出时，

$$\max \frac{\frac{1}{m}\sum_{i=1}^{m}(s_i^-/g_i^x)+\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}(s_r^+/g_r^y)+\sum_{t=1}^{q_2}(s_t^{b-}/g_t^b)\right)}{2}$$

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

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

$$\mathrm{B}\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda,s^-,s^+\geq 0 \qquad (4.65)$$

在方向 SBM 模型的目标函数中，投入指标数量 m 和产出指标数量 q（或 q_1+q_2）并没有像 RAM 和 BAM 模型那样合并在一起。因此，在方向 SBM 模型中如果将非期望产出作为投入处理，得到的结果是不同的。

方向 SBM 模型得出的结果是无效率值，如果是计算面板数据的 TFP 指数，则只能采用 Luenberger 指数方法（加减运算）。那么如何才能得出效率值并采用 Malmquist 指数(乘除运算)呢？

下面来看一下方向 SBM 模型与 RAM 模型和 BAM 模型的关系。

方向 SBM 模型与 RAM 模型和 BAM 模型三者本质上都属于加权加性模型，三个模型都可以看做是采用了不同权重的加权加性模型，RAM 和 BAM 是在加权加性模型的基础上，将结果中

的无效率值转换为效率值（转换方法：效率值 = 1 - 无效率值）。

方向 SBM 模型得出的无效率值=$\max \sum_{i=1}^{m}\frac{1}{2mg_i^x}s_i^- + \sum_{r=1}^{q}\frac{1}{2qg_r^y}s_r^+$

RAM 模型得出的效率值为 1-无效率值

$$=\min\ 1-\left(\sum_{i=1}^{m}\frac{s_i^-}{(m+q)R_i^-}+\sum_{r=1}^{q}\frac{s_r^+}{(m+q)R_r^+}\right)$$

$$=\min\ 1-\left(\sum_{i=1}^{m}\frac{s_i^-}{2\mathrm{m}(m+q)R_i^-/(2m)}+\sum_{r=1}^{q}\frac{s_r^+}{2\mathrm{q}(m+q)R_r^+/(2q)}\right)$$

RAM 模型与方向 SBM 模型的关系：RAM 模型可以看作是方向 SBM 模型的一个特例，即 RAM 模型相当于伪方向向量 $g_i^x=\frac{(m+q)R_i^-}{2m}$，$g_r^y=\frac{(m+q)R_r^+}{2q}$，并采用“效率值 =1-无效率值”方法进行转换的方向 SBM 模型。

BAM 模型与方向 SBM 模型的关系：BAM 模型可以看作是方向 SBM 模型的一个特例，即 BAM 模型相当于伪方向向量 $g_i^x=\frac{(m+q)L_{ik}^-}{2m}$，$g_r^y=\frac{(m+q)U_{rk}^+}{2q}$，并采用“效率值 = 1-无效率值”方法进行转换的方向 SBM 模型。

如果既想使用“方向 SBM 模型”这一名称，又要得出“效率值”（可计算 Malmquist 指数），则可以选择 RAM 模型或 BAM 模型，因为 RAM 模型和 BAM 模型可以看作是将伪方向向量设置为上述特定数值，并将无效率值转换为效率值的“方向 SBM 模型”。

除此之外，也可以在后期采用自定义的方法将软件输出的方向 SBM 模型的无效率值转换为效率值，并重新手工计算 Malmquist 指数（效率值转换方法必须保证得出的效率值为正数，否则无法计算 Malmquist 指数）。

4.6.7 加权加性距离之非径向方向距离函数(non-radial directional distance function)

这是一种在文献中应用较多的非径向方向距离函数，参见文献(Barros, Managi, & Matousek, 2012; Zhang, Zhou, & Choi, 2013; Mirsalehy, Abu Bakar, Lee, Jaafar, &

Heydar, 2014; Yao, Zhou, Zhang, & Li, 2015; Lee & Choi, 2018)。

其表达式为

$$\max \sum_{i=1}^{m}\omega_i^x\beta_i^x+\sum_{r=1}^{q_1}\omega_r^y\beta_r^y+\sum_{t=1}^{q_2}\omega_t^b\beta_t^b$$

$$s.t.\ \sum_{j=1}^{n}x_{ij}\lambda_j+\beta_i^x g_i^x\le x_{ik},i=1,2,3,...,m$$

$$\sum_{j=1}^{n}y_{rj}\lambda_j-\beta_r^y g_r^y\ge y_{rk},r=1,2,3...,q_1$$

$$\sum_{j=1}^{n}b_{tj}\lambda_j+\beta_t^b g_t^b=b_{tk},t=1,2,3...,q_2$$

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

$$\lambda,\beta^x,\beta^y,\beta^b\ge 0 \qquad (4.66)$$

（$-g^x,g^y,-g^b$）是伪方向向量。

在此将模型中所谓的方向向量称为伪方向向量，是因为模型中并不存在真正的方向向量，只是形似而已。上述模型并不能像真正的方向距离函数模型那样采用简洁的矩阵和向量表达式。

注意模型(4.66)中的不等号改为等号后，模型结果保持不变[i]，即

[i] 以其中第 m 项投入的约束为例来说明在此模型中不等号与等号等价。

$\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x\le x_{m0}$ 等价于 $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x+\sigma_m=x_{m0}$，$\sigma_m\ge 0$。由于目标函数是最大化各项 β 加权之和，假设在模型一个解中存在 $\sigma_m^*>0$，则一定可以在保持 $\beta_i\ \ (i\ne m)$ 的最优解 $\beta_i^*\ \ (i\ne m)$ 不变的条件下，得到一个更优解 $\beta_m^{*2}=\beta_m^*+\sigma_m^*,\sigma_m^{*2}=0$。因此，在模型的最优解中，必然存在 $\sigma_m^*=0$。从而得出结论 $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x\le x_{m0}$ 等价于 $\sum_{j=1}^{n}x_{mj}\lambda_j+\beta_m^x g_m^x=x_{m0}$。

$$\max \sum_{i=1}^{m}\omega_i^x\beta_i^x+\sum_{r=1}^{q_1}\omega_r^y\beta_r^y+\sum_{t=1}^{q_2}\omega_t^b\beta_t^b$$

$$s.t.\ \sum_{j=1}^{n}x_{ij}\lambda_j+\beta_i^x g_i^x=x_{ik},i=1,2,3,...,m$$

$$\sum_{j=1}^{n}y_{rj}\lambda_j-\beta_r^y g_r^y=y_{rk},r=1,2,3...,q_1$$

$$\sum_{j=1}^{n}b_{ij}\lambda_j+\beta_t^b g_t^b=b_{tk},t=1,2,3...,q_2$$

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

$$\lambda,\beta^x,\beta^y,\beta^b\geq 0 \qquad (4.67)$$

令 s = βg，w = ω/g，则模型(4.67)可转化为

$$\max \sum_{i=1}^{m}w_i^-s_i^-+\sum_{r=1}^{q_1}w_r^+s_r^++\sum_{t=1}^{q_2}w_t^{b-}s_t^{b-}$$

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

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

$$B\lambda+s^{b-}=b_k$$

$$e\lambda=1$$

$$\lambda,s^-,s^+\geq 0 \qquad (4.68)$$

转换后的模型与包含了非期望产出的加权加性模型(4.55)完全相同。

在非径向方向距离函数模型中，如果将非期望产出作为投入处理，得到的结果是相同的。

可以看出，上述非径向方向距离函数模型就是权重为 w = ω/g 的加权加性模型。

注意非径向方向距离函数的伪方向向量为（$-g^x,g^y,-g^b$）。 g^x,g^y,g^b 都是正数，因此得出的权重（w^x,w^y,w^b）也都是正数。

上述转换过程要求伪方向向量值 g > 0。当一项投入或产出的伪方向向量值 g 需要设置为 0 时，该项投入或产出在目标函数中对应的 w = 0，对应的约束如下：

如果是投入，则 $\sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i^x g_i^x \le x_{ik}$，得出 $\sum_{j=1}^{n} x_{ij}\lambda_j \le x_{ik}$，即 $X\lambda + s^- = x_k$。与模型(4.68)中的投入约束项相同。

如果是产出，则 $\sum_{j=1}^{n} y_{rj}\lambda_j - \beta_r^y g_r^y \ge y_{rk}$，得出 $\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rk}$，即 $Y\lambda - s^+ = y_k$。与模型(4.68)中的产出约束项相同。

如果是非期望产出，则 $\sum_{j=1}^{n} b_{tj}\lambda_j + \beta_t^b g_t^b = b_{tk}$，得出 $\sum_{j=1}^{n} b_{tj}\lambda_j = b_{tk}$，相当于 $B\lambda + s^{b-} = b_k$，$s^{b-} = 0$。与模型(4.68)中的非期望产出约束项的差别是增加了 $s^{b-} = 0$ 的限制[i]。

在 MaxDEA 软件中，只需选择加权加性模型下的“自定义” 分项（h 或 i 分项），即可实现上述非径向方向距离函数。

h) 自定义（所有 DMU 使用相同权重）

i) 自定义（各 DMU 使用不同权重）

首先按照上面的公式计算出权重，即 $w^x = \omega^x / g^x$，$w^y = \omega^y / g^y$，$w^b = \omega^b / g^b$。

然后再设置权重即可。h 是直接填写计算出的权重，i 是将计算出的权重作为数据中的一个权重变量。

使用 h 分项时，直接为每项投入产出填写权重数值（软件默认的权重数值均为 1），设置的权重适用于所有 DMU；

使用 i 分项时，为每项投入产出选择权重变量，因此可以为不同 DMU 设置不同的权重。使用 i 分项时，必须在导入数据之前，在数据中为每项投入和产出增加相应的权重变量。注意权重变量在定义数据时，保持“Not defined”不变。

在 MaxDEA 软件的输出结果中，提供了 Inefficiency Score (无效率值)和投入产出的松弛改进（slack movement）。如果用户需要各变量的 β 值，可按照以下公式计算：

$$\beta_i = s_i^* / g_i \qquad (4.69)$$

s_i^* 是结果中各项投入产出的松弛改进（slack movement）的绝对值，g_i 是用户定义的各

[i] s=0，在 MaxDEA 软件中对应的设置为：高级模型选项选择“不可随意控制的投入产出”，并设置该项非期望产出的“松弛变量自由程度”为 0。

项投入产出的方向向量值。公式中 $s_i^* \geqslant 0$，$g_i > 0$。

以表 4-25 为例，以$(-x_0, y_0, -b_0)$作为伪方向向量，两项投入(X1 和 X2)、期望产出(Y)和非期望产出(B)的权重(ω)分别为 0.25，0.25，0.3，0.2。

表 4-25 非径向方向距离函数示例数据

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

各项指标的最终权重(w)为 $w^x = \omega^x / g^x$，$w^y = \omega^y / g^y$，$w^b = \omega^b / g^b$，即 $w^{x1} = 0.25 / x_1$，$w^{x2} = 0.25 / x_2$，$w^y = 0.3 / y$，$w^b = 0.2 / b$。

将计算出的 4 项权重作为变量增加到数据中。在定义数据时保持“Not defined”。权重类型选择 “i) 自定义 (各 DMU 使用不同权重)”，并设置各项指标的权重变量；规模报酬选择 VRS；在“高级模型”界面，将 B 设置为非期望产出。

表 4-26 是分析结果及根据公式(4.69)计算得出的 Beta 值。读者可以通过 β 值和权重 ω (0.25, 0.25, 0.3, 0.2) 来计算无效率值，即无效率值 = $\sum_{i=1}^{m} \omega_i^x \beta_i^x + \sum_{r=1}^{q_1} \omega_r^y \beta_r^y + \sum_{t=1}^{q_2} \omega_t^b \beta_t^b$，验证一下是否与软件给出的无效率值一致。

表 4-26 非径向方向距离函数分析结果

DMU	无效率值	松弛改进[i]				Beta 值			
		X1	X2	Y	B	X1	X2	Y	B

[i] 在 MaxDEA 软件的结果中，松弛改进采用正负号来表示改进的方向，正数表示增加，负数表示减少。

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

非径向方向距离函数得出的结果是无效率值，如果是计算面板数据的 TFP 指数，只能采用 Luenberger 指数方法（加减运算）。那么如何才能得出效率值并采用 Malmquist 指数(乘除运算)呢？与上一节中提出的方法相同。

下面来看一下非径向方向距离函数模型与 RAM 模型和 BAM 模型的关系。

非径向方向距离函数模型与 RAM 模型和 BAM 模型三者本质上都属于加权加性模型，三个模型都可以看做是采用了不同权重的加权加性模型，RAM 和 BAM 是在加权加性模型的基础上，将结果中的无效率值转换为效率值（转换方法：效率值 = 1 - 无效率值）。

非径向方向距离函数模型得出的无效率值=$\max \sum_{i=1}^{m}\frac{\omega_i^-}{g_i^-}s_i^- + \sum_{r=1}^{q_1}\frac{\omega_r^+}{g_r^+}s_r^+$

RAM 模型得出的效率值为 1-无效率值

$$=\min\ 1-\left(\sum_{i=1}^{m}\frac{s_i^-}{(m+q)R_i^-}+\sum_{r=1}^{q}\frac{s_r^+}{(m+q)R_r^+}\right)$$

$$=\min\ 1-\left(\sum_{i=1}^{m}\frac{\omega_i^-}{(m+q)R_i^-/\omega_i^-}s_i^-+\sum_{r=1}^{q}\frac{\omega_r^+}{(m+q)R_r^+/\omega_r^+}s_r^+\right)$$

RAM 模型与非径向方向距离函数模型的关系：RAM 模型可以看作是非径向方向距离函数的一个特例，即 RAM 模型相当于伪方向向量 $g_i^x=\frac{(m+q)R_i^-}{\omega_i^-}$，$g_r^y=\frac{(m+q)R_r^+}{\omega_r^+}$，并采用“效

率值 = 1-无效率值”方法进行转换的非径向方向距离函数模型。

BAM 模型与非径向方向距离函数模型的关系：BAM 模型可以看作是非径向方向距离函数的一个特例，即 BAM 模型相当于伪方向向量 $g_i^x = \frac{(m+q)L_{ik}^-}{\omega_i^-}$，$g_r^y = \frac{(m+q)U_{rk}^+}{\omega_r^+}$，并采用“效率值 = 1-无效率值”方法进行转换的非径向方向距离函数模型。

如果既想使用“非径向方向距离函数”这一名称，又要得出“效率值”（可计算 Malmquist 指数），则可以选择 RAM 模型或 BAM 模型，因为 RAM 模型和 BAM 模型可以看作是将伪方向向量设置为上述特定数值，并将无效率值转换为效率值的“非径向方向距离函数”。

除此之外，也可以在后期采用自定义的方法将软件输出的非径向方向距离函数的无效率值转换为效率值，并重新手工计算 Malmquist 指数（效率值转换方法必须保证得出的效率值为正数，否则无法计算 Malmquist 指数）。

4.6.8 各类加权加性距离之间的关系

各类加权加性距离本质上属于同一类距离。

1）简单加性距离与方向 SBM 模型的关系

简单加性距离可以看作是方向 SBM 模型的一个特例，即相当于伪方向向量 $g_i^x = \frac{1}{2m}$，$g_r^y = \frac{1}{2q}$ 的方向 SBM 模型。

2）简单加性距离与非径向方向距离函数模型的关系

简单加性距离可以看作是非径向方向距离函数的一个特例，即相当于伪方向向量 $g_i^x = \omega_i^-$，$g_r^y = \omega_r^+$ 的非径向方向距离函数模型。

3）RAM 模型与方向 SBM 模型的关系

RAM 模型可以看作是方向 SBM 模型的一个特例，即 RAM 模型相当于伪方向向量 $g_i^x = \frac{(m+q)R_i^-}{2m}$，$g_r^y = \frac{(m+q)R_r^+}{2q}$，并采用“效率值 = 1-无效率值”方法进行转换的方向 SBM 模型。

4）RAM 模型与非径向方向距离函数模型的关系

RAM 模型可以看作是非径向方向距离函数的一个特例，即 RAM 模型相当于伪方向向量

$g_i^x = \frac{(m+q)R_i^-}{\omega_i^-}$，$g_r^y = \frac{(m+q)R_r^+}{\omega_r^+}$，并采用“效率值 =1-无效率值”方法进行转换的非径向方向距离函数模型。

5）BAM 模型与方向 SBM 模型的关系

BAM 模型可以看作是方向 SBM 模型的一个特例，即 BAM 模型相当于伪方向向量 $g_i^x = \frac{(m+q)L_{ik}^-}{2m}$，$g_r^y = \frac{(m+q)U_{rk}^+}{2q}$，并采用“效率值 = 1-无效率值”方法进行转换的方向 SBM 模型。

6）BAM 模型与非径向方向距离函数模型的关系

BAM 模型可以看作是非径向方向距离函数的一个特例，即 BAM 模型相当于伪方向向量 $g_i^x = \frac{(m+q)L_{ik}^-}{\omega_i^-}$，$g_r^y = \frac{(m+q)U_{rk}^+}{\omega_r^+}$，并采用“效率值 = 1-无效率值”方法进行转换的非径向方向距离函数模型。

4.7 EBM 混合距离

EBM 混合距离函数是由 K. Tone & Tsutsui (2010)提出的一种包含径向与 SBM 两类距离函数的混合模型，因模型中使用 ε 参数，Tone 将其称为 epsilon-based measure(EBM)。

4.7.1 EBM 模型

下面以投入导向 CRS 模型为例，介绍 EBM 模型的计算过程。投入导向 CRS EBM 模型的线性规划式表示为

$$\min\ \theta-\varepsilon\frac{1}{\sum_{i=1}^{\mathrm{m}}w_i^-}\sum_{i=1}^{m}\frac{w_i^-s_i^-}{x_{\mathrm{k}}}$$

$$s.t.\ X\lambda-\theta x_k+s^-=0$$

$$Y\lambda\geq y_{\mathrm{k}}$$

$$\lambda\geq 0, s^-\geq 0 \qquad (4.70)$$

被评价 DMU 的效率值为目标函数的最优解，即

$\theta^*-\varepsilon\sum_{i=1}^{m}\frac{w_i^-s_i^{-*}}{x_{\mathrm{k}}}$。

模型中有 m+1 个参数：ε 和 w_i^- (i=1,2,…,m)。w_i^-表示各项投入指标的相对重要程度，ε 是一个关键参数，取值范围为[0, 1]，它表示在效率值的计算中，非径向部分的重要程度：取 0 时相当于径向模型；取 1 时相当于 SBM 模型。

这些参数值需要在建立 EBM 模型之前确定。K. Tone & Tsutsui (2010)认为这些参数应由数据本身来确定，并提出了计算方法。以两项投入（x1 和 x2）的模型为例，方法如下：

首先，通过 SBM 模型或其它模型获得各项投入指标的投影值[i]。

此处为 X1 和 X2 两项投入指标，其投影值记为 P1 和 P2。两个指标投影值之间的关联性代表着在该项生产技术（生产前沿）中两项投入的数量比例关系。通过分析二者的数量关系，可以得出它们之间在生产过程中的可替代性：如果两种投入的数量呈现高度线性正相关，说明二者之间可替代性差，需要以相对固定的比例进行生产，对生产效率的测量以径向测量

[i] 在 MaxDEA 软件中，投影值的计算方法采用 SBM 模型。

为主，ε 宜取较小数值，甚至为 0；如果两种投入的数量呈现高度线性负相关，则说明二者之间可替代性强，不要求以固定比例进行生产，对生产效率的测量以非径向模型为主，ε 宜取较大数值，甚至为 1。

然后，建立投入指标投影值的关联矩阵，矩阵的元素由各项投入指标投影值两两之间的关联指数（affinity index）构成。

此处为两项投入，该矩阵表示为：

表 4-27 关联指数矩阵

	x1	x2
x1	S(P1,P1)	S(P1,P2)
x2	S(P2,P1)	S(P2,P2)

S( )是计算两个投入指标的投影值 P1 和 P2 之间关联指数的函数。K. Tone & Tsutsui (2010)认为关联指数应符合以下规则：

(1) 同一性（Identical）: S(a,a) = 1，自身关联指数＝1；

(2) 对称性（Symmetric）: S(a,b) = S(b,a)，计算的指数与两个指标的前后顺序无关；

(3) 单位不变性（Units-invariant）: S(ta,b) = S(a,b) (t > 0)，数据的测量单位对计算的指标数值没有影响；

(4) 数值范围：0≤S(a,b) ≤ 1，计算的指数数值在 0-1 之间。

遵循这样的思路，我们首先想到的是 S( )采用线性相关系数，即 Pearson 相关系数。K. Tone & Tsutsui (2010)认为，Pearson 相关系数未能满足上述条件。

K. Tone & Tsutsui (2010)的方法是先用离散指数函数计算指标两两之间的离散指数。然后用离散指数计算关联指数：

S(a,b)=1-2D(a,b)。

离散指数计算方法为

$$D(a,b)=\begin{cases}\dfrac{\sum_{j=1}^{n}|c_j-\bar{c}|}{n(c_{\max}-c_{\min})} & (if\ c_{\max}>c_{\min}),\\ 0 & (if\ c_{\max}=c_{\min})\end{cases} \qquad (4.71)$$

$c_j=\ln\frac{b_j}{a_j}$，$\bar{c}=\frac{1}{n}\sum_{j=1}^{n}\ln\frac{b_j}{a_j}$，$c_{\max}=\max(c_j)$, and $c_{\min}=\min(c_j)$。

最后，利用建立的关联指数矩阵，计算 EBM 中所需的参数。

$$\varepsilon = \frac{m - \max(\rho)}{m-1}$$

$$w_{\mathrm{i}} = \frac{v_i}{\sum_{i=1}^{m} v_i}$$

ρ 是关联指数矩阵的最大特征根，v 是其对应的特征向量。

产出导向 EBM 模型的规划式表示为

$$\min \frac{1}{\phi + \varepsilon^{+} \frac{1}{\sum_{r=1}^{q} w_r^{+}} \sum_{r=1}^{q} \frac{w_r^{+} s_r^{+}}{y_{\mathrm{k}}}}$$

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

$$Y\lambda - \phi y_{\mathrm{k}} - s^{+} = 0$$

$$\lambda \ge 0, s^{-} \ge 0 \qquad (4.72)$$

非导向 EBM 模型的规划式表示为

$$\min \frac{\theta - \varepsilon^{-} \frac{1}{\sum_{i=1}^{\mathrm{m}} w_i^{-}} \sum_{i=1}^{m} \frac{w_i^{-} s_i^{-}}{x_{\mathrm{k}}}}{\phi + \varepsilon^{+} \frac{1}{\sum_{r=1}^{q} w_r^{+}} \sum_{r=1}^{q} \frac{w_r^{+} s_r^{+}}{y_{\mathrm{k}}}}$$

$$s.t.\ X\lambda - \theta x_k + s^{-} = 0$$

$$Y\lambda - \phi y_{\mathrm{k}} - s^{+} = 0$$

$$\lambda \ge 0, s^{-} \ge 0 \qquad (4.73)$$

4.7.2 EBM 模型的缺陷与改进方法

问题 1：EBM 模型的投影值会出现逻辑错误。

在投入导向 EBM 模型(4.70)中，变量 θ 取值范围没有限制，K. Tone & Tsutsui (2010)对此也没有讨论。在对变量 θ 取值范围没有约束的情况下，模型最优解 θ^{*}有可能大于 1。无效 DMU 的投入指标的投影值有可能高于其原始值，即会出现改进目标是“增加投入”的错误。

同样，在产出导向 EBM 模型(4.72)中，对变量 φ 取值范围没有约束的情况下，模型最

优解 φ^*有可能小于 1。无效 DMU 的产出指标的投影值有可能低于其原始值，即会出现改进目标是“减少产出”的错误。

在非导向 EBM 模型(4.73)中，问题更为突出，上述两种错误都可能会出现。

在此以表 4-21 数据为例演示这一问题，采用 VRS 模型，分别求解投入导向、产出导向和非导向 EBM 模型，EBM 参数方法采用 K. Tone & Tsutsui (2010)的关联指数方法。错误结果见表 4-28、表 4-29、表 4-30。

为了避免产生这样的错误，需要在 EBM 模型中对 θ 和 φ 进行约束，以非导向 EBM 模型为例，修正后的规划式为

$$\min \frac{\theta-\varepsilon^{-}\dfrac{1}{\sum_{i=1}^{\mathrm{m}} w_i^{-}}\sum_{i=1}^{m}\dfrac{w_i^{-}s_i^{-}}{x_{\mathrm{k}}}}{\phi+\varepsilon^{+}\dfrac{1}{\sum_{r=1}^{q} w_r^{+}}\sum_{r=1}^{q}\dfrac{w_r^{+}s_r^{+}}{y_{\mathrm{k}}}}$$

$$s.t.\ X\lambda-\theta x_k+s^{-}=0$$

$$Y\lambda-\phi y_{\mathrm{k}}-s^{+}=0$$

$$\lambda\geq 0, s^{-}\geq 0, \theta\leq 1, \phi\geq 1 \qquad (4.74)$$

EBM 模型与径向和 SBM 模型的关系：

1）如果 ε＝0，则 EBM 模型等同于径向模型。

2）如果 ε＝1，则 EBM 模型等价于（加权）SBM 模型，此时由于增加了变量 θ 和/或 φ，模型可能会存在无穷多个最优解，但所有最优解的效率值和投影值均相等。

表 4-28 EBM 模型错误结果（投入导向 VRS）

地区	效率值	θ	床位数原始值	床位数投影值
广西	0.9834	1.0001	95752	95760
江西	0.9979	1.0037	87184	87504
天津	0.9855	1.0295	40787	41991

表 4-29 EBM 模型错误结果（产出导向 VRS）

地区	效率值	φ	入院人数 原始值	入院人数 投影值
安徽	0.8414	0.9929	4391516	4360170
广西	0.9217	0.9749	3258438	3176512
贵州	0.8257	0.8953	2433370	2178564
河南	0.8412	0.9580	7043628	6747705
湖北	0.8907	0.9984	4805124	4797549
湖南	0.8263	0.9408	5393968	5074549
江苏	0.9520	0.9949	6433837	6401057
江西	0.8780	0.9097	3000964	2729925
山东	0.8318	0.9370	8354234	7827682
四川	0.9196	0.9199	6489314	5969408
新疆	0.8625	0.8795	3115413	2740144

表 4-30 EBM 模型错误结果（非导向 VRS）

地区	效率值	θ	φ	床位数原始值	床位数投影值	卫技人员数原始值	卫技人员数投影值	诊疗人次数原始值	诊疗人次数投影值	入院人数原始值	入院人数投影值
安徽	0.8407	**1.1771**	1.1718	140997	165965	132739	156245				
甘肃	0.8048	**2.0742**	2.4323	66661	138266	53127	110195				
广西	0.9163	**2.2404**	2.1091	95752	214524	105773	236975				
贵州	0.8209	**1.9629**	1.9375	78368	153830	69320	136070				
海南	0.7464	**11.5154**	13.0687	21367	246050	26329	289388				
河北	0.7959	0.9346	**0.9649**							5675067	5475952
河南	0.8112	0.7144	**0.7576**							7043628	5336206
黑龙江	0.6273	**1.3622**	1.8426	129449	176338	127358	173489				
湖北	0.8903	**1.1378**	1.1231	152062	173022	147628	167977				
湖南	0.8232	0.6101	**0.5290**							5393968	2853139
吉林	0.6934	**1.6842**	2.1743	94636	159390	86278	145313				
江苏	0.9095	0.6091	**0.6697**					166944454	111801185	6433837	4308682
江西	0.8685	**2.3167**	2.1414	87184	201978	93287	216116				
内蒙古	0.6604	**2.5478**	3.3446	72871	185659	74177	188987				
宁夏	0.7982	**8.0824**	9.2620	22037	178112	21830	176439				

问题 2：关联指数计算方法错误，不能反映真实关联程度，甚至得出完全相反的结果。

K. Tone & Tsutsui (2010)指出，关联指数的计算具有以下特点：两个指标的投影值呈现极度径向分布的情况下，关联指数应该等于 1；而在极度离散的情况下，关联指数应该等于 0（图 4-11）。K. Tone & Tsutsui (2010) 给出了两个简单的示例来说明这一特点，其中极度离散的示例（该文献中提供的 Example 2），计算得出的关联指数为 0。但是，这一示例过于简单，只有两个有效 DMU，不能说明问题。

在此给出一个简单的极度离散的示例(表 4-31)，来说明 K. Tone & Tsutsui (2010)提出的关联指数计算方法并不符合上述要求。如图 4-11 所示，两项投入指标投影值之间呈现完全线性负相关的关系[i]。

[i] 无效 DMU 的投影值为 5 个有效 DMU 其中之一。

表 4-31 关联指数计算示例数据(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

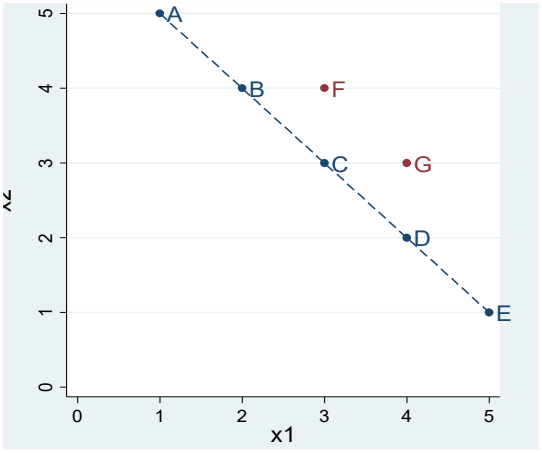


图 4-11 离散数据示例

根据前面关联指数的计算方法，得出的关联指数为 S(P1,P2)=0.428, ε＝0.572。这一结果与 K. Tone & Tsutsui (2010)建立关联指数的理论相悖。

表 4-32 示例数据的离散指数矩阵

	x1	x2
x1	0	0.286
x2	0.286	0

表 4-33 示例数据的关联指数矩阵

	x1	x2
x1	1	0.428
x2	0.428	1

上面例子说明，这种关联指数计算方法不能反映真实关联程度。下面再看两个得出完全错误结果的例子。

表 4-34 中 x1 和 x2 是两个几乎完全关联的变量（

图 4-12），误差仅为 0.0001%。但是采用 K. Tone & Tsutsui (2010)的方法计算得出的离散指数 D(x1, x2) = 0.5，关联指数 S(x1, x2) = 0。

表 4-35 中 x1 和 x2 是两个高度关联的变量（图 4-13）。但是采用 K. Tone & Tsutsui (2010)

的方法同样得出离散指数 D(x1, x2) = 0.5，关联指数 S(x1, x2) = 0。

以上两个示例说明，K. Tone & Tsutsui (2010)计算离散指数和关联指数的方法是错误的，其离散指数计算公式(4.71)反映的并非两个变量的离散程度，而是平均离散值（$\frac{\sum_{j=1}^{n}|c_j-\bar{c}|}{n}$）与离散范围（$c_{\max}-c_{\min}$）的比值。上面两个示例说明，只要平均离散值是离散范围的一半，不管离散范围有多小（只要不等于 0），得出的离散指数 D(x1, x2)总是等于其可能的最大值 0.5，进而得出关联指数 S(x1, x2) = 0。只要平均离散值接近离散范围的一半，得出的离散指数 D(x1, x2)接近 0.5，关联指数 S(x1, x2) 接近于 0。

表 4-34 关联数据示例 (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

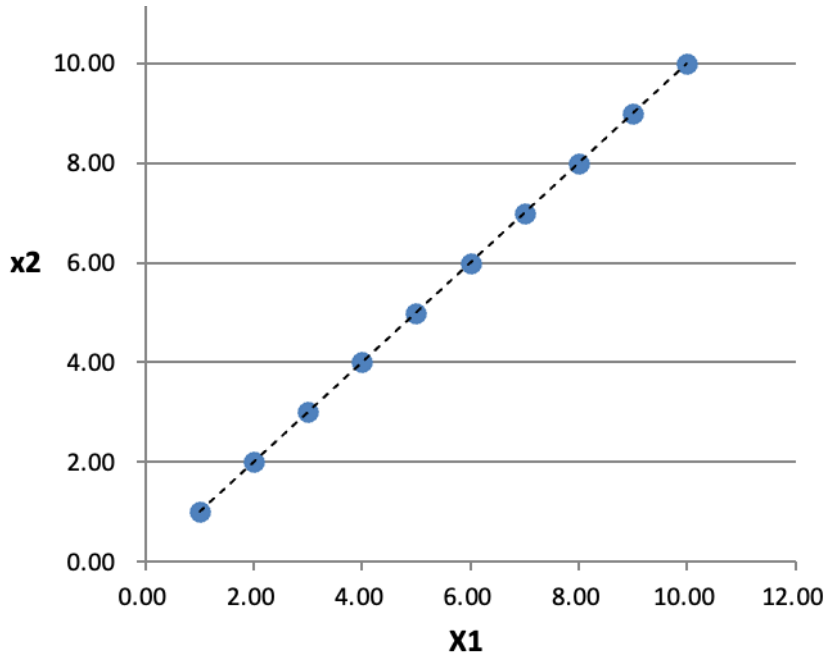


图 4-12 关联数据示例(1)

表 4-35 关联数据示例 (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

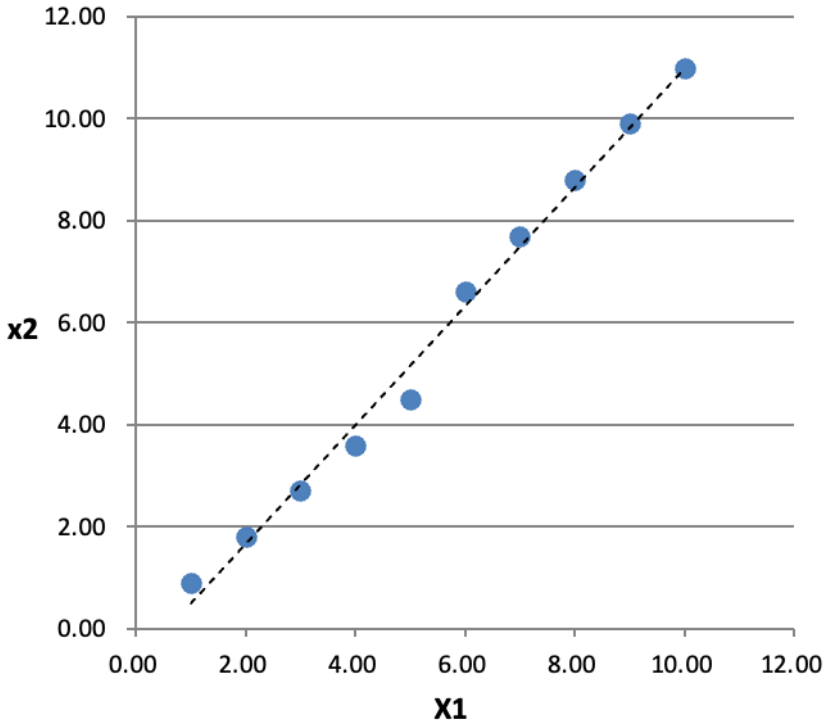


图 4-13 关联数据示例(2)

我们再回到最常用来表示两个指标关联程度的 Pearson 相关系数。事实上，Pearson 相关系数满足 K. Tone & Tsutsui (2010)提出的第 1～3 条规则，只是不满足第 4 条规则。但是只需通过简单的数据转换，即可将其数值范围调整为[0, 1]。

S(a,b)=0.5+0.5R(x1,x2),

R(x1,x2) 是两个指标的 Pearson 相关系数。

如果两项指标的投影值呈现径向分布，其 Pearson 相关系数数值较大，则说明两种投入之间可替代性差；如果两项指标的投影值呈现类似图 4-11 那样的离散分布，其 Pearson 相关系数数值为负数，且绝对值较大，两种投入之间的可替代性强。

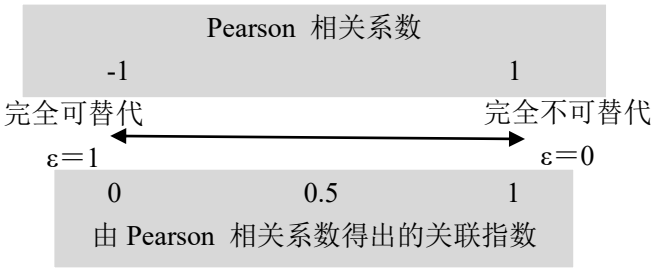


图 4-14 Pearson 相关系数与新的关联指数的关系

由 表 4-31 中数据计算得出的新关联指数

S(P1,P2)=0.5+0.5 R(P1, P2)=0.5+0.5(-1)=0。

表 4-36 新关联指数矩阵

DMU	x1	x2
x1	1	0
x2	0	1

新关联指数矩阵的最大特征根为 1，因此

$$\varepsilon = \frac{m-\max(\rho)}{m-1} = \frac{2-1}{2-1} = 1,$$

此时 EBM 模型等价于 SBM 模型。

虽然经过以上分析，采用由 Pearson 相关系数计算得出的关联指数来构建 EBM 模型的参数更为适宜，但是在实际应用时，EBM 模型的参数 ε 和 w 完全可以由决策者根据生产技术本身的特性和决策偏好来决定，而不是由数据自身来确定。

4.7.3 MaxDEA 软件求解混合距离函数模型操作方法

在 MaxDEA 软件中，混合距离函数模型的求解除 Distance 的设置之外，其它选项与其它距离函数类型相同。

EBM 模型需要选择生成参数(ε 和 w)的方法：

1）采用 K. Tone & Tsutsui (2010)的方法；

2）采用本书中介绍的调整 Pearson 相关系数；

3）由用户输入参数，包括 ε 和 w(指标权重)，投入和产出的 ε 值可直接填写（Coefficient for Input Nonradial Measure, Coefficient for Output Nonradial Measure），点击“设置”按钮，打开投入和产出指标权重设置界面。

4.8 包含价格信息的距离

4.8.1 成本效率、收益效率模型

DEA 得到广泛应用的原因之一是技术效率的分析不需要投入和产出指标的价格信息。但是，当价格信息已知时，则可以进一步利用价格信息对配置效率进行分析。例如，如果投入要素的价格已知，则可以分析成本效率，并通过比较成本效率与技术效率，得出配置效率(Farrell, 1957; R. Färe, Grosskopf, & Lovell, 1985)。被评价 DMU k 的成本效率（cost efficiency, CE）模型表示为

$$\text{Cost Efficiency} = \frac{cx^*}{cx_k}$$

$$cx^* = \min \ cx$$

$$s.t. \ x \geq X\lambda$$

$$y_k \leq Y\lambda$$

$$\lambda \geq 0 \qquad (4.75)$$

c 表示价格向量。成本效率、配置效率(allocative efficiency)和技术效率的关系为

成本效率（CE）＝技术效率（TE）×配置效率（AE）。

当产出的价格信息已知时，可以分析收益效率（revenue efficiency）：

$$\text{Revenue Efficiency} = \frac{ry_k}{ry^*}$$

$$ry^* = \max \ ry$$

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

$$y \leq Y\lambda$$

$$\lambda \geq 0 \qquad (4.76)$$

当投入和产出的价格信息均已知时，还可以分析利润效率(profit efficiency)：

表 4-37 成本、收益效率模型的规划式

成本效率	$\text{Cost Efficiency} = \frac{cx^*}{cx_k}$ $cx^* = \min \ cx$ $s.t. \ x \geq X\lambda$ $y_k \leq Y\lambda$ $\lambda \geq 0$
收益效率	$\text{Revenue Efficiency} = \frac{ry_k}{ry^*}$ $ry^* = \max \ ry$ $s.t. \ x_k \geq X\lambda$ $y \leq Y\lambda$ $\lambda \geq 0$

c 表示投入价格向量，r 表示产出价格向量

4.8.2 MaxDEA 软件求解成本效率等模型的操作方法

求解成本效率等模型，除了投入和产出指标之外，在数据表中还要增加相应的价格指标。

在将数据导入 MaxDEA 软件后，在数据定义时，价格指标保持“Not defined”不变。求解成本效率模型，必须设置投入的价格；求解收益效率模型，必须设置产出的价格。

模型距离选择成本或收益，然后点击右侧的设置按钮，从下拉框中选择投入和/或产出指标的价格变量。成本效率等模型不存在模型导向的选择问题，但可以选择规模报酬类型。

MaxDEA 软件在结果中没有提供配置效率，用户可根据需要进行计算，例如分别运行投入导向径向 DEA 模型和成本模型，然后计算得到的成本效率与技术效率之比值即为配置效率。

4.9 单个投入产出指标无效率程度分析

在 DEA 效率值和投影分析的基础上，可以进一步对单个投入产出指标的无效率程度进行分析。

DMU k 的单个投入指标的无效率程度计算方法为

$\frac{x_{ik}-\hat{x}_{ik}}{x_{ik}}$ ，$\hat{x}_{ik}$ 为 x_{ik} 的投影值；

单个产出指标的无效率程度计算方法为

$\frac{\hat{y}_{rk}-y_{rk}}{y_{rk}}$ ，$\hat{y}_{rk}$ 为 y_{rk} 的投影值。

需要注意的是，上述单个投入产出指标无效率程度仅仅是针对当前的效率测量距离类型及其投影点而言的。不同的距离类型会得出不同的投影，从而得出不同的单个投入产出指标无效率程度。即使是同一距离类型，例如方向距离函数，当采用不同的方向向量时也会得到不同的投影。例如下图中的方向距离函数，当采用不同的方向向量时，G 的投影点位于折线 ABCD 上。当投影为 A 点时，投入 X_2 的无效率程度为 0，投入的无效率全部集中于 X_1；当投影为 D 点时，投入 X_1 的无效率程度为 0，投入的无效率全部集中于 X_2。因此，单个投入产出指标无效率程度分析结果不能脱离所采用的效率测量方法及其投影点。不能简单地说某个 DMU 的某项投入或产出指标的无效率程度等于多少，这样是没有意义的。单个投入产出指标无效率程度分析应视为是对所选择的效率测量方法及其投影的进一步解释，并非是对被评价 DMU 生产过程某项投入（产出）“浪费”（“不足”）程度的评价。

MaxDEA 不提供单个投入产出指标无效率程度分析结果，用户可以利用软件给出的投影值自己计算。

第5章 超效率模型

在 DEA 模型的分析结果中，通常会出现多个 DMU 被评价为有效的情况，尤其是当投入和产出指标数量较多时，有效 DMU 数量也会较多。DEA 模型得出的效率值最大为 1，有效 DMU 效率值相同，这些有效 DMU 的效率高低无法进一步区分。为了解决这一问题，Andersen & Petersen (1993)提出了能够对有效 DMU 进一步区分其有效程度的方法，这一方法后来被称为“超效率”模型(super-efficiency model)。为便于表述，把前面介绍的模型称为标准效率模型（standard efficiency model）。超效率模型的核心就是将被评价 DMU 从参考集中剔除，也就是说，被评价 DMU 的效率是参考其它 DMU 构成的前沿得出的，有效 DMU 的超效率值一般会大于 1，从而可以对有效 DMU 进行区分。

在许多效率分析应用中，需要对效率的影响因素做进一步分析。由于效率值最大为 1（即存在上限），被视为截尾数据，所以在文献中多采用 Tobit 回归模型。超效率则不存在效率值的截尾问题，因此无需采用专门的处理截尾数值的 Tobit 回归模型。

虽然 Andersen & Petersen (1993)提出的是径向超效率模型，但其原理同样适用于其它类型的距离函数。例如 K. Tone (2002)提出的 SBM 超效率模型，Ray (2008)提出的方向距离函数超效率模型等。

对于超效率模型的排序功能，Banker & Chang (2006)通过数据模拟发现其对有效 DMU 的排序结果并不可靠，但是这并未影响超效率模型得到越来越多的应用。

超效率模型其它常用功能包括：

1）发现异常数据(Wilson, 1995; Banker & Chang, 2006)；

2）DEA 模型的敏感性和稳定性分析(W. W. Cooper et al., 2001; J. Zhu, 2001; Jahanshahloo, Hosseinzadeh, Shoja, Sanei, & Tohidi, 2005; Boljuncic, 2006; F. Hosseinzadeh Lotfi, Jahanshahloo, & Esmaeili, 2007)；

3）广义 DEA（后面专门介绍）；

4）通过跨期比较，分析生产率的变动情况（后面专门介绍）。

5.1 径向超效率模型

以投入导向 CRS 径向模型为例，超效率模型与标准效率模型的唯一区别就是增加了 $j\neq k$ 这一限制条件，即从参考集中剔除被评价 DMU k(Andersen & Petersen, 1993)：

$$\min\theta$$

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

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

$$\lambda\ge 0$$

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j=1,2\ldots,n\ (j\neq k) \qquad (5.1)$$

VRS 超效率模型就是在 CRS 模型的基础上增加约束 $\sum_{j=1,j\neq k}^{n}\lambda_j=1$：

$$\min\theta$$

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

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

$$\sum_{\substack{j=1\\ j\neq k}}^{n}\lambda_j=1$$

$$\lambda\ge 0$$

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j=1,2\ldots,n\ (j\neq k) \qquad (5.2)$$

下面以图示的方式来说明超效率模型的基本原理。同样以投入导向径向标准效率模型图示数据表 2-3 为例，在标准效率模型中，A、B、C、D 四个有效 DMU 构成有效前沿。现在以分析 C 的超效率为例，C 的超效率是参考由除 C 以外的其它 DMU 构成的前沿，即 ABD，C 在该前沿上的投影点为 C'。[i] 显然 C 的效率优于 C'，C 与其投影点 C'相比，其效率超出的部分反映为 CC'，其超效率值 θ^*＝OC'/ OC≥1。

对于有效 DMU 来说，在投入导向的超效率模型中，其径向投影方向是增加投入。在保持其它 DMU 投入和产出数值不变的条件下，该有效 DMU 在等比例增加投入后（增加比例为 θ^*-1），仍将被评价为有效，这是超效率模型用于敏感性和稳定性分析的基本原理。

[i] 注意超效率模型的投影点也可能位于标准效率模型的前沿上，例如假如在前沿上 AB 两点之间还有一个 DMU，则该 DMU 的超效率投影点为其自身，仍然位于标准效率模型的前沿上。

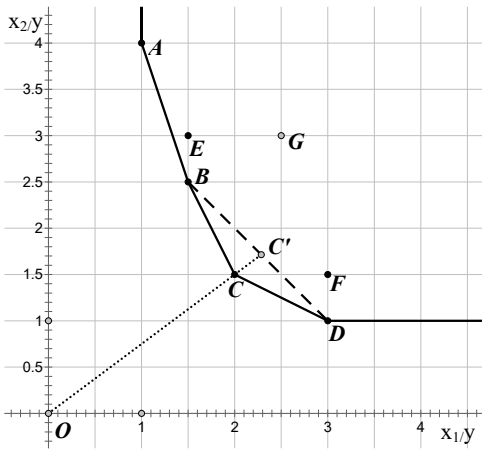


图 5-1 超效率模型示意图

产出导向 VRS 超效率模型表示为

$$\max \phi$$

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

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

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j = 1$$

$$\lambda \ge 0$$

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.3)$$

虽然超效率模型是针对有效 DMU 而设计，但同样适用于无效 DMU，无效 DMU 的超效率径向模型与标准效率径向模型的结果相同。

5.2 方向距离函数超效率模型

与径向超效率模型的原理相同，方向距离函数超效率模型也是在标准效率模型的基础上增加 j≠k 这一限制条件(Ray, 2008)。以 VRS 模型为例，包含非期望产出的方向距离函数超效率模型为：

$$\max \ \beta$$

$$\text{s.t.} \sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j x_{ij} + \beta g_{\text{xi}} \le x_{ik}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j y_{rj} - \beta g_{yr} \ge y_{rk}$$

$$\sum_{\substack{j=1 \\ j\neq k}}^{n} \lambda_j b_{tj} - \beta g_{yt} \le \text{b}_{tk}$$

$$\sum_{j=1, j\neq k}^{n} \lambda_j = 1$$

$$\lambda \ge 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q_1; t = 1,2\ldots,q_2; j = 1,2\ldots,n \ (j \neq k) \qquad (5.4)$$

与径向超效率模型相同，方向距离函数超效率模型也同样适用于无效 DMU，无效 DMU 的超效率模型与标准效率模型的结果相同。

5.3 SBM(ERM)超效率模型

与径向和方向距离函数超效率模型相比，SBM(ERM)超效率模型要复杂一点，它并非仅仅是增加 j≠k 这一限制条件(K. Tone, 2002)。SBM(ERM)超效率的规划式仅适用于有效 DMU，对于 SBM(ERM)有效 DMU k，以 VRS 为例，其非导向 SBM(ERM)超效率模型表示为

$$\min \rho_{\mathrm{SE}} = \frac{\frac{1}{m}\sum_{i=1}^{m} \bar{x}_i \ / x_{ik}}{\frac{1}{q}\sum_{r=1}^{q} \bar{y}_r \ / y_{rk}}$$

$$s.t. \ \bar{x}_i \geq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j$$

$$\bar{y}_r \leq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j$$

$$\bar{x}_i \geq x_{ik}$$

$$\bar{y}_r \leq y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

$$\lambda, s^-, s^+, \bar{y} \geq 0$$

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n \ (j \neq k) \quad (5.5)$$

由除 DMU k 之外的其它 DMU 构建的生产可能集为

$\{(x,y): x \geq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, y \leq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j\}$。

被评价 DMU k 在 SBM(ERM)超效率模型中的投影值 $(\bar{x}, \bar{y})$，即模型的最优解为在由其它 DMU 构建的生产可能集内，距离前沿最近的点。为什么是距离前沿最近的点，而不是一定位于前沿上呢？

为了便于与 SBM(ERM)标准效率模型进行比较，我们对模型(5.5)进行如下等价变换：

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

$$s.t.\ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

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

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.6)$$

为什么在 SBM(ERM)标准效率模型中，约束使用的是等号，而在超效率模型中使用的是不等号呢？

SBM(ERM)超效率模型得到的是被评价 DMU 距离由其它 DMU 构成的前沿的最小距离，因此 SBM(ERM)超效率模型的投影点有可能位于前沿的弱有效部分，在这种情况下，如果在 SBM(ERM)超效率模型的约束中使用等号，有可能会造成模型无可行解。

在此以 VRS SBM(ERM)超效率模型为例，如果在约束中使用等号，则模型为

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

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

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

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

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

$$i=1,2\ldots,m; r=1,2\ldots,q; j=1,2\ldots,n\ (j\neq k) \quad (5.7)$$

令 $\hat{x}=\sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^-$，$\hat{y}=\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+$，则 $(\hat{x},\hat{y})$ 的可行域为

$$\{(\hat{x},\hat{y}): \hat{x}\leq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, \hat{y}\geq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j, \sum_{j=1,\, j\neq k}^{n} \lambda_j = 1\}$$。

如果 DMU k 的某项投入大于其它任何 DMU，或者某项产出小于任何其它 DMU，即

$x_{ik} > x_{ij}, j\neq k$ 或 $y_{rk} < y_{rj}, j\neq k$，

则 $(x_k, y_k)\notin \{(\hat{x},\hat{y}): \hat{x}\leq \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j, \hat{y}\geq \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j, \sum_{j=1,\, j\neq k}^{n} \lambda_j = 1\}$，即 DMU k 不在 $(\hat{x},\hat{y})$ 的可行域范围内。

这是因为在 $\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$ 约束下，由 $x_{ik} > x_{ij}, j\neq k$ 或 $y_{rk} < y_{rj}, j\neq k$ 必然推导出

$$x_{ik} > \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j \text{ 或 } y_{ik} < \sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j$$。

这说明如果 $x_{ik} > x_{ij}, j\neq k$ 或 $y_{rk} < y_{rj}, j\neq k$，则模型(5.7)无可行解。

以单投入单产出的 VRS 模型为例，表 5-1 的数据分布如图 5-2 所示。C 的投入 x 数值比其它所有 DMU 都高，如果采用模型(5.7)计算 C 的超效率，$(\hat{x},\hat{y})$ 的可行域为图中阴影区域，C 不在该区域内，所以模型(5.7)无解。A 的产出 y 数值低于其它所有 DMU，如果采用模型(5.7)计算 A 的超效率，也会无可行解。

表 5-1 SBM(ERM)超效率模型示例数据

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

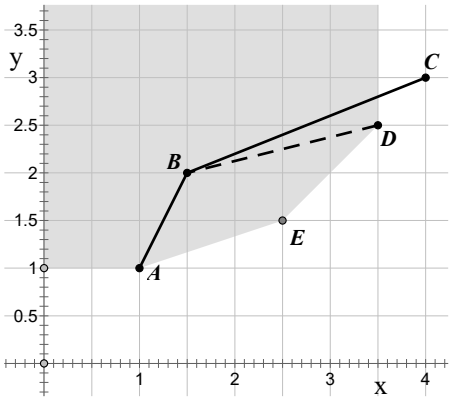


图 5-2 模型(5.7)出现无可行解示意图

虽然以上讨论基于 VRS 模型，但无解的问题不仅限于 VRS 模型，在 CRS 模型中也会出现，但其原因不像 VRS 模型这样易于解释。例如，在 CRS SBM(ERM)超效率模型中，如果约束使用等号，表 2-3 中的 A 无可行解[i]。

SBM(ERM)超效率模型(5.6)为非导向 SBM(ERM)超效率模型，投入导向和产出导向的 SBM(ERM)超效率模型分别表示为

$$\min \rho_{\mathrm{SE}} = 1 + \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}$$

[i] 注意在下一节专门介绍 VRS 超效率模型的无解问题，但该问题与现在讨论的不是同一个问题。后面介绍的 VRS 超效率模型的无解问题不仅存在于投入导向和产出导向的 SBM 超效率模型中，同样也存在于投入导向和产出导向的径向超效率模型和方向距离函数超效率模型中。

$$s.t.\ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

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

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

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n\ (j \neq k) \quad (5.8)$$

$$\min \rho_{\mathrm{SE}} = \frac{1}{1 - \frac{1}{q}\sum_{r=1}^{q} s_r^+ / y_{rk}}$$

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

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

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

$$i = 1,2\ldots,m; r = 1,2\ldots,q; j = 1,2\ldots,n\ (j \neq k) \quad (5.9)$$

超效率 SBM(ERM)模型得到的是被评价 DMU 距离由其它 DMU 构成的前沿的最小距离，其投影点并非一定是强有效的，因此超效率 SBM(ERM)模型可认为也是 MinDS 和 MinDW 模型的超效率求解模型。

包含非期望产出时的 SBM(ERM)超效率模型（VRS）为

$$\min \rho_{\mathrm{SE}}=\frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1} s_r^+ / y_{rk}+\sum_{t=1}^{q_2} s_t^{b-} / \mathrm{b}_{tk}\right)}$$

$$s.t.\ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} b_{tj}\lambda_j - s_t^{b-} \le b_{tk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

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

$$i=1,2\ldots,m;\ r=1,2\ldots,q_1;\ t=1,2\ldots,q_2;\ j=1,2\ldots,n\ (j\neq k) \quad (5.10)$$

5.4 加权加性距离超效率模型

加权加性距离超效率模型与 SBM(ERM)超效率模型相似，不再详述。

以 VRS 模型为例，如果 DMU k 为有效 DMU（即无效率值=0），则其加权加性距离超效率模型为

$$\max \ -\sum_{i=1}^{m} w_i^- s_i^- - \sum_{r=1}^{q} w_r^+ s_r^+$$

$$s.t. \ \sum_{j=1,\, j\neq k}^{n} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j=1,\, j\neq k}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j b_{tj} - s_t^{b-} \le \mathrm{b}_{tk}$$

$$\sum_{j=1,\, j\neq k}^{n} \lambda_j = 1$$

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

$$i=1,2\ldots,m; r=1,2\ldots,q_1; t=1,2\ldots,q_2; j=1,2\ldots,n\ (j\neq k) \quad (5.11)$$

加权加性距离超效率模型得出的无效率值为负数。

5.5 超效率模型的投影分析

本书第 2 章（2.12 对投影值容易产生的误解）解释了 DEA 模型投影值的含义。投影值（目标值）是被评价的 DMU 在前沿上的投影，一个 DMU 的当前值与投影值之间的相对差距代表着该 DMU 各项投入产出的无效率程度。

在标准效率模型中，一个无效的 DMU，其投影值总是优于其原始值，即投入投影值≤原始值，产出投影值≥原始值，非期望产出投影值≤原始值。

在超效率模型中，无效 DMU 的结果（包括投影值）与标准效率模型中的结果是完全相同的。而有效的 DMU（效率值≥1），其投影值和原始值之间不存在固定的大小关系。超效率模型中，被评价的有效 DMU 是往由其他 DMU 所构成的前沿上做投影，被评价 DMU 处于该前沿的前面，其投影值与原始值之间的关系比较复杂，这与模型的距离和导向类型有关。这里提供一个简单的总结表，有兴趣的读者可以进一步去分析。

表 5-2 超效率模型中有效 DMU（效率值≥1）的投影值与原始值的关系

距离类型	投入导向	产出导向	非导向
径向距离和方向距离函数[i]	投入投影值与原始值的关系不确定（通常是投入投影值≥原始值） 好产出投影值≥原始值 坏产出投影值≤原始值	投入投影值≤原始值 产出投影值与原始值的关系不确定（通常是好产出投影值≤原始值，坏产出投影值≥原始值）	投入投影值与原始值的关系不确定（通常是投入投影值≥原始值） 产出投影值与原始值的关系不确定（通常是好产出投影值≤原始值，坏产出投影值≥原始值）
至前沿的最远距离（ERM, SBM）	投入投影值≥原始值 好产出投影值≥原始值 坏产出投影值≤原始值	投入投影值≤原始值 好产出投影值≤原始值 坏产出投影值≥原始值	投入投影值≥原始值 好产出投影值≤原始值 坏产出投影值≥原始值
加权加性距离			投入投影值≥原始值 好产出投影值≤原始值 坏产出投影值≥原始值

[i] 以非导向超效率模型为例，投影值与目标值之间的差距由两部分构成，一是比例变化部分，二是松弛变化部分。二者符号（变化方向）相反。投影值与目标值之间的大小关系取决于比例变化部分和松弛部分哪个绝对值更大。

5.6 通过超效率模型探测异常数据

DEA 的非参数前沿本质决定了，如果一个异常数据出现在前沿上，则有可能会出现大量异常结果，甚至是所有其它 DMU 的分析结果都会异常。因此，分析时确认是否存在异常值是一项重要工作。当数据量较大时，通过直接观察发现异常值非常困难。

超效率模型的一项重要功能就是发现异常值，下面通过将表 2-6 中的数据修改为异常数值，然后通过超效率模型来识别出异常值。将福建的投入数据减少至原来的 10%，其它数据保持不变。改变前后的结果比较见表 5-3。福建的数值改变后，在投入导向模型的分析结果中，几乎所有 DMU 的效率值都发生了改变。特别是 CRS 模型，变化尤为突出，除福建自身外，其它所有 DMU 的效率值都降低至原来的 1/10 至 1/6；VRS 的变化相对较小，有 4 个有效 DMU 的结果没有受到影响，其它 DMU 的效率值降低幅度呈现连续性变化，影响最小的变化幅度仅 4%，最大的降低至原来的接近 1/10。从超效率看，福建 CRS 和 VRS 超效率值均大于 10，与其数值变化的比例相符。

从这一简单的示例可以看出，超效率模型是一种有效的发现异常数据的方法。在实际应用中，可以通过改变距离函数、模型导向和 RTS 类型，分析多种超效率模型来探测可能的异常数据。

表 5-3 通过超效率模型探测异常值（投入导向）

	CRS			VRS		
地区	**正常效率值**	**异常数据效率值**	**异常数据超效率值**	**正常效率值**	**异常数据效率值**	**异常数据超效率值**
安徽	0.9442	0.1006	0.1006	0.9549	0.5786	0.5786
北京	0.9174	0.1487	0.1487	0.9185	0.4967	0.4967
福建	1.0000	**1.0000**	**10.4553**	**1.0000**	**1.0000**	**10.5133**
甘肃	0.8324	0.0979	0.0979	0.8538	0.1458	0.1458
广东	1.0000	0.1490	0.1490	1.0000	1.0000	1.3051

第6章 其他模型与方法

6.1 乘数比值约束模型

在乘数模型中，如果对投入和产出权重的比值增加额外的约束，则称为乘数比值约束模型（restricted multiplier model），又称为锥比率模型、约束锥模型、保证域模型(assurance region)。

6.1.1 对两项投入的权重比值进行约束

对一项投入和另一项投入的权重的比值设置上下限的约束表示为

$$R_L \le \frac{v_a}{v_b} \le R_U \qquad (6.1)$$

该比值的经济学意义为投入 a 对投入 b 的边际技术替代率(marginal rate of technical substitution，MRTS)，即在保持产量不变的条件下，增加一种生产要素 a 的数量，与可以减少的另一种生产要素 b 的数量之比。在等产量曲线上，该比值表示一个点的切线的斜率。由于生产过程中投入的生产要素是不完全替代的，随着一种生产要素数量的增加，该要素对另外一种要素的边际技术替代率是递减的。

当研究者知道两项投入之间的边际替代数量关系时，增加这样的约束可以使分析结果更为合理。例如通过增加权重比值约束可以解决分析结果中出现权重为 0 的问题（权重为 0 的问题详见第 2 章 2.13.3）。当研究者缺乏对这种数量关系的了解时，可以将比值约束的上下限设置为研究样本分析结果中相应权重比值的最大和最小值。例如在图 6-1 中，可以观察到，前沿由 ABCD 四个顶点构成，H 的投影点 H'位于 A 点与横坐标的平行线上（弱有效前沿），因此得出的 X_1 的权重必然为 0。投入 X_1 和 X_2 的权重之比(v_1/v_2)的最大值存在于 CD 段，比值（即斜率的绝对值）为 2；权重比值的最小值存在于 AB 段，比值为 0.5。如果将 X_1 和 X_2 的权重之比限制在 0.5-2.0 之间，即

$$0.5 \le \frac{v_1}{v_2} \le 2,$$

H 的投影点将变为 H"，从而有效解决 X_1 的权重为 0 及投影点为弱有效的问题。

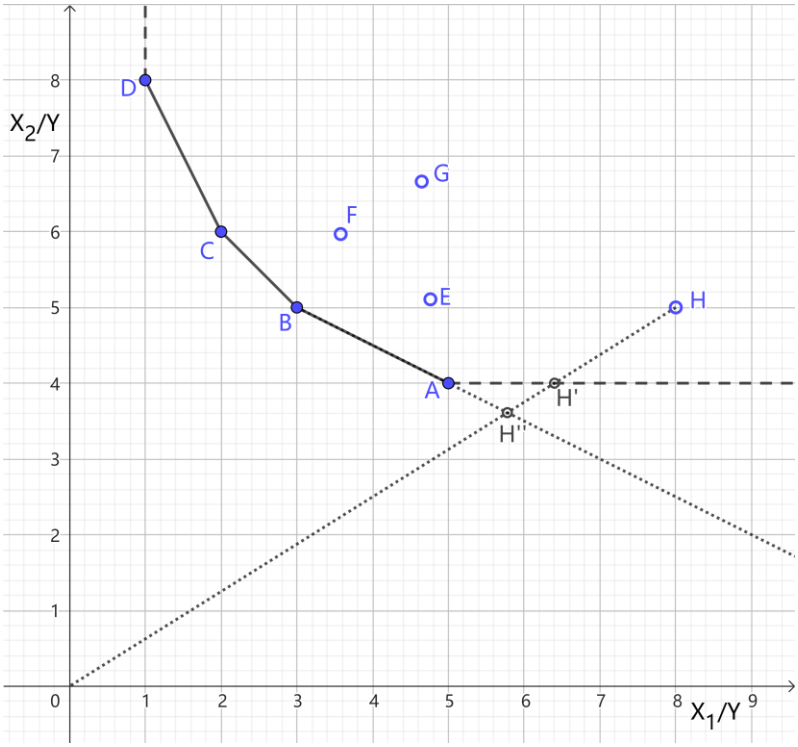


图 6-1 权重比值约束示例（2 投入 1 产出 CRS 投入导向模型）

6.1.2 对一项产出和一项投入的权重比值进行约束

$$R_L \le \frac{v_a}{u_b} \le R_U \qquad (6.2)$$

当只有一项投入和一项产出时，该比值的经济学意义为 x 的边际生产率(marginal productivity)，即产量增加的数量与投入要素增加的数量之比，在生产前沿曲线上，表示切线的斜率或投入 x 的导数。当有多项投入和多项产出时，该比值是该项投入对该项产出的偏导数，其含义为在其他投入和产出保持不变的条件下，该项投入对该项产出的边际生产率。在图 6-2 中，可以观察到，前沿由 ABCD 四个顶点构成，H 的投影点 H'位于 A 点与纵坐标的平行线上（弱有效前沿），因此得出的 Y 的权重必然为 0。投入 X 和产出 Y 的权重之比(v/u)的最大值存在于 AB 段，比值（即斜率）为 2；权重比值的最小值存在于 CD 段，比值为 0.5。如果将投入 X 和 Y 的权重之比限制在 0.5-2.0 之间，即

$$0.5 \le \frac{v}{u} \le 2,$$

H 的投影点将变为 H"，从而有效解决 Y 的权重为 0 及投影点为弱有效的问题。

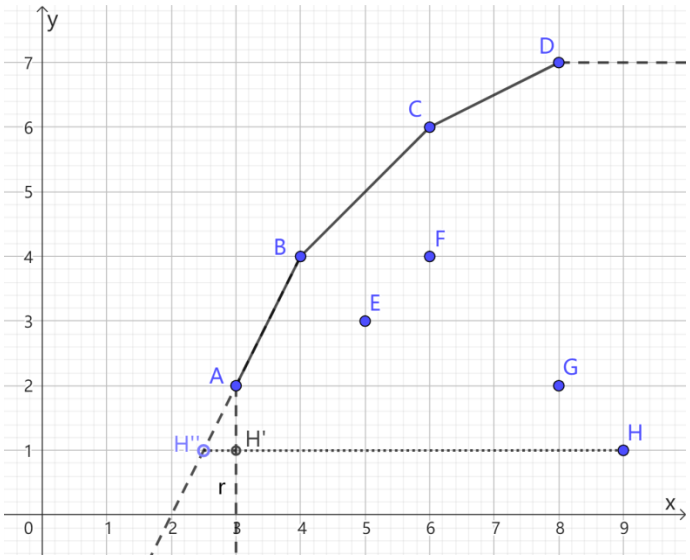


图 6-2 权重比值约束示例（1 投入 1 产出 VRS 投入导向模型）

6.1.3 对一项产出和另一项产出的权重比值进行约束

对一项产出和另一项产出的权重的比值设置上下限的约束表示为

$$R_L \le \frac{u_a}{u_b} \le R_U \qquad (6.3)$$

其经济学含义类似边际技术替代率，在保持投入数量不变的条件下，减少一种产品的数量，与可以增加的另一种产品的数量之比。设置产出权重比值约束同样可以解决结果中出现权重为 0 的问题，在此不再重复。

6.2 FDH 模型

FDH 模型是由 Tulkens (1993)提出的一种混合整数线性规划（mixed integer linear programming, MILP）模型，从规划式上看，FDH 模型可以看作是在 VRS DEA 模型的基础上，将线性组合系数 λ 限定为要么是 0，要么是 1，即 $\lambda \in \{0,1\}$ 。FDH 模型的全称是 Free Disposal Hull，直接含义为自由处置壳（包）。FDH 这一名称主要体现的是 Free disposal 的含义，即在生产可能集中投入和产出的可自由处置性（强可处置）。

以投入导向模型为例，FDH 模型表示为

$$\min \theta$$

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

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

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

$$\lambda \in \{0,1\}$$

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j=1,2\ldots,n \qquad (6.4)$$

由于模型中约束$\lambda \in \{0,1\}$， FDH 模型的前沿不再是凸包，其外形如图 6-3 所示，虚线为 VRS DEA 模型前沿，实线为 FDH 前沿。

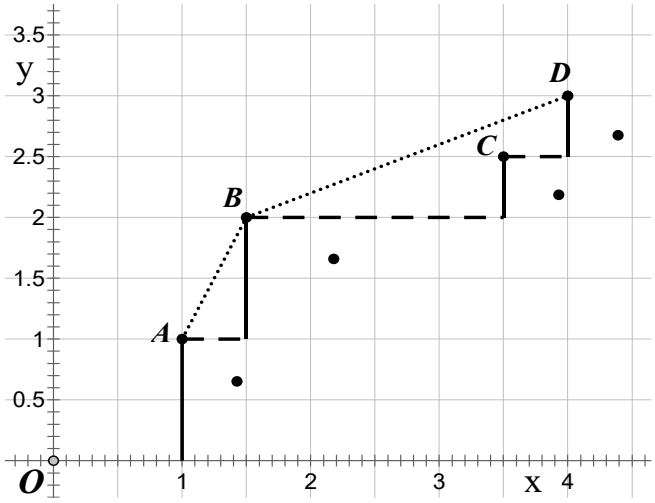


图 6-3 FDH 模型示意图（投入导向）

从图中可以看出，FDH 模型的生产可能集是 VRS DEA 模型的子集。在 FDH 模型中，被评价 DMU 的参考标杆只能有一个 DMU，并且其线性组合系数 λ＝1。其意义在于被评价 DMU 仅参考现实中实际存在的 DMU，不会参考线性组合而成的虚拟 DMU，这是 FDH 模型在实际应用中的意义。

FDH 模型可以从上述投入导向的径向模型，拓展到产出导向、非导向模型，以及其他距离类型。

6.3 包含非期望产出的模型

关于在 DEA 模型中如何处理非期望产出指标，在前面的方向距离函数部分已作了详细介绍。对包含非期望产出 DEA 模型的结果的理解，详见本书“4.5.8 对包含非期望产出 DEA 模型的结果容易产生的误解”。

在此再总结一下非期望产出在各类距离函数中的处理方法。

6.3.1 包含非期望产出的径向模型

$$\begin{aligned}&\max\ \beta\\&s.t.\ X\lambda+\beta x_k\le x_k\\&\quad Y\lambda-\beta y_k\ge y_k\\&\quad B\lambda-\beta b_k\le b_k\\&\quad \lambda\ge 0\end{aligned}\qquad (6.5)$$

6.3.2 包含非期望产出的 SBM 模型

Tone 定义的包含非期望产出的 SBM 模型为(William W Cooper et al., 2007)：

$$\min\ \rho=\frac{1-\frac{1}{m}\sum_{i=1}^{m}s_i^-/x_{ik}}{1+\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}s_r^+/y_{rk}+\sum_{t=1}^{q_2}s_t^{b-}/\mathrm{b}_{tk}\right)}$$

$$\begin{aligned}&s.t.\ X\lambda+s^-=x_k\\&\quad Y\lambda-s^+=y_k\\&\quad \mathrm{B}\lambda+s^{b-}=b_k\\&\quad \lambda,s^-,s^+\ge 0\end{aligned}\qquad (6.6)$$

本书在前面的讨论中建议在方向距离函数或径向模型中，将非期望产出设置为强可处置性。从 SBM 的规划式可以反映出其生产可能集正是基于非期望产出强可处置性的。

Tone 没有给出包含非期望产出的 SBM 超效率模型的规划式，但根据 SBM 超效率模型**(5.6)**可推出包含非期望产出的 SBM 超效率模型为

$$\min\ \rho=\frac{1+\frac{1}{m}\sum_{i=1}^{m}s_i^-/x_{ik}}{1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}s_r^+/y_{rk}+\sum_{t=1}^{q_2}s_t^{b-}/\mathrm{b}_{tk}\right)}$$

$$s.t.\ \sum_{j=1,\,j\neq k}^{n}x_{ij}\lambda_j-s_i^-\le x_{ik}$$

$$\sum_{j=1,\,j\neq k}^{n}y_{rj}\lambda_j+s_r^+\ge y_{rk}$$

$$\sum_{j=1,\,j\neq k}^{n}b_{tj}\lambda_j-s_t^{b-}\le b_{tk}$$

$$1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}s_r^+/y_{rk}+\sum_{t=1}^{q_2}s_t^{b-}/\mathrm{b}_{tk}\right)>0$$

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

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j=1,2\ldots,n\ (j\neq k)\quad (6.7)$$

其中增加的约束$1-\frac{1}{q_1+q_2}\left(\sum_{r=1}^{q_1}s_r^+/y_{rk}+\sum_{t=1}^{q_2}s_t^{b-}/\mathrm{b}_{tk}\right)>0$可在线性转换时去除。

在 DEA 模型中，还可通过将非期望产出表示为负数的方式来处理。另外，在 MaxDEA 软件中，没有提供专门的非期望投入模型，可通过将非期望投入指标转换为负数的方式解决，详见后面负数指标的介绍。

MaxDEA 软件支持负数指标，并支持各种模型选项的组合应用，例如包含非期望产出的 SBM 超效率模型，只需选择 SBM 模型，并在 Advanced Models(1)界面勾选 SuperEfficiency 和 Undesirable Outputs 并进行相应设置即可。

6.3.3 包含非期望产出的方向距离函数模型

$$\max \beta$$

$$s.t.\ X\lambda+\beta g_x \le x_k$$

$$Y\lambda-\beta g_y \ge y_k$$

$$B\lambda-\beta g_b \le b_k$$

$$\lambda \ge 0$$

$$g_x \ge 0, g_y \ge 0, g_b \le 0 \quad (6.8)$$

6.3.4 包含非期望产出的加权加性模型

6.3.4.1 包含非期望产出的简单加性模型

$$\max \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{q_1} s_r^+ + \sum_{t=1}^{q_2} s_t^{b-}$$

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

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

$$\mathrm{B}\lambda+s^{b-} = b_k$$

$$\lambda, s^-, s^+ \ge 0 \quad (6.9)$$

6.3.4.2 包含非期望产出的加权加性模型

$$\max \sum_{i=1}^{m} w_i^- s_i^- + \sum_{r=1}^{q_1} w_r^+ s_r^+ + \sum_{t=1}^{q_2} w_t^{b-} s_t^{b-}$$

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

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

$$\mathrm{B}\lambda+s^{b-} = b_k$$

$$\lambda, s^-, s^+ \ge 0 \quad (6.10)$$

6.3.5 通过负数指标实现对非期望产出的处理

部分 DEA 模型允许投入和产出指标中存在负数，例如自定义方向向量的方向距离函数模型，RDM 方向距离函数模型，MSBM 模型。

径向模型、SBM 和 MinDS 模型：投入导向模型允许产出数据中有负数；产出导向模型允许投入数据中有负数；非导向模型中不允许有负数[i]。

以径向模型为例来说明为什么不允许有负数。投入导向径向 VRS 模型可等价表示为

$$\min 1-\beta$$

$$s.t.\ X\lambda+\beta x_k \le x_k$$

$$Y\lambda \ge y_k$$

$$e\lambda=1$$

$$\lambda \ge 0 \qquad (6.11)$$

$x_{ik}-\alpha x_{ik}$ 是被评价 DMU 的投入指标在径向改进后的目标值，改进值为 $-\alpha x_{ik}$，负数表示改进方向为减少投入。当投入指标 x_k 为负数时，改进值变为正数，径向改进方向变成了等比例增加投入。

同理，在产出导向模型中，当产出指标 y_k 为负数时，也会出现改进方向变为等比例减少产出的错误结果。

在径向模型中如果将改进部分 $-\alpha x_k$ 中的 x_k 取绝对值，就可以避免出现上述逻辑错误，并且如果投入指标为正数，修改后的模型与修改前的模型等价(Cheng, Zervopoulos, & Qian, 2013)。修改后的径向模型（variant of radial measure, VRM）表示为

$$\min 1-\beta$$

$$s.t.\ X\lambda+\beta|x_k| \le x_k$$

$$Y\lambda \ge y_k$$

$$e\lambda=1$$

$$\lambda \ge 0 \qquad (6.12)$$

在上述公式中| |表示变量的数值取绝对值。修改后的 DEA 线性规划式，在不包含负数时，即 x_k 为正数时，其最优解与传统的 DEA 线性规划式等价，也就是说，新的 DEA 模型保持了与传统 DEA 模型的兼容性。当 y_k 为负数时，取绝对值保证了无效率改进方向的正确

[i] 有学者认为，上述允许有负数的情况仅适用于 VRS 模型，因为只有 VRS 模型才符合平移不变性（translation-invariance）：包含负数的指标同时加上一个正数之后，模型的结果保持不变，从而可以将负数指标通过平移转换为正数指标。笔者认为，平移不变性是允许负数的充分条件，而非必要条件。上述允许有负数的情况可否用于其他 RTS 类型，决定于所研究的技术和数据是否符合相应的 RTS 类型特征。如果技术符合投入变为原来的 λ 倍（Xλ），产出包括负产出会相应的变为原来的 λ 倍（该技术符合 CRS 特征），则上述允许有负数的情况可应用于 CRS 模型（例如用负数表示非期望产出）。

性。

修改后的产出导向 VRM 模型表示为

$$\min \ \frac{1}{1+\beta}$$

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

$$Y\lambda - \beta |y_k| \ge y_k$$

$$e\lambda = 1$$

$$\lambda \ge 0 \qquad (6.13)$$

同理可以对 SBM 模型进行类似处理，但只需对 SBM 模型的目标函数进行修改即可：

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

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

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

$$e\lambda = 1$$

$$\lambda, s^-, s^+ \ge 0 \qquad (6.14)$$

VRM 模型存在的两个问题：

1）当投入或产出指标中既有负数，也有正数时，修改后模型得出的效率值不具备单调性，即有可能出现被评价 DMU 减少投入后或增加产出后，得出的效率值反而下降的问题。在这种情况下，可将一个既包含正数也包含负数的指标分解为两个指标，分别保存指标的正数部分和负数部分：[i]

$$v_{ij}^1 = \begin{cases} v_{ij} & if\ v_{ij} \ge 0, \\ 0 & if\ v_{ij} < 0, \end{cases} \text{ 和 } v_{ij}^2 = \begin{cases} 0 & if\ v_{ij} \ge 0, \\ v_{ij} & if\ v_{ij} < 0. \end{cases} \qquad (6.15)$$

[i] 这一分离正负数的方法是借鉴 Emrouznejad, Anouze, & Thanassoulis (2010)提出的 SORM（semi-oriented radial measure）方法，但有区别。Emrouznejad et al. (2010)的方法分离后的两个变量保存的均为正数，即 $v_{ij}^1 = \begin{cases} v_{ij} & if\ v_{ij} \ge 0, \\ 0 & if\ v_{ij} < 0, \end{cases}$ 和 $v_{ij}^2 = \begin{cases} 0 & if\ v_{ij} \ge 0, \\ -v_{ij} & if\ v_{ij} < 0. \end{cases}$。

2）VRM 模型需有条件应用于 CRS。CRS VRM 模型仅适用于以下情况：负数指标中只包含负数，且伴随着生产规模扩大（投入增加），负数指标的绝对值增加。例如将非期望产出用负数表示，随着生产规模的扩大，非期望产出也会相应增加。

如果用负数表示非期望产出，VRM 模型与前文介绍的包含非期望产出的径向模型等价。包含负数产出指标的 VRM 模型（产出导向）为

$$\min \ \frac{1}{1+\beta}$$

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

$$Y\lambda - \beta |y_k| \ge y_k$$

$$\lambda \ge 0 \qquad (6.16)$$

其中，约束$Y\lambda - \beta |y_k| \ge y_k$可分解为正数产出指标和负数产出指标的约束：

$Y^+\lambda - \beta y_k \ge y_k^+$，

$Y^-\lambda - \beta(-y_k) \ge y_k^-$。

负数产出指标约束的两边同乘以-1，

$(-Y^-)\lambda + \beta(-y_k) \le (-y_k^-)$，

与包含非期望产出的产出导向径向模型**(4.42)**等价。

类似地，可通过负数指标来处理非期望投入问题，例如垃圾和废水处理等。

6.4 MaxDEA 软件中对负数的处理

部分 DEA 模型本身支持负数，例如方向距离函数、加权加性距离、MSBM 模型。当选择的模型本身不支持负数时，MaxDEA 软件会自动应用上述 VRM（或类似处理方法）以使选择的模型允许负数。

6.5 外部不可控因素或不可控投入/产出指标

在效率或生产率分析的实际应用中，经常会遇到生产过程受到某些决策者所不能控制的外部因素影响的情况，例如公共卫生服务效率会受到服务区域内交通便利程度、人口密度等因素的影响。如果这些因素能够归结为数值变量，并且转换为正向指标，即值越大说明生产环境越优，则可以在 DEA 模型中将这类外部影响因素作为不可随意控制（uncontrollable,

non-discretionary）的投入指标处理。另外，在生产实践中也存在着部分投入（产出）数量不是决策者所能决定或控制的情况，这种情况下，同样可作为不可随意控制的投入（产出）指标处理。例如医院提供的医疗服务数量不是医院决策者所能直接决定的，但可以通过提高服务质量等方式吸引患者就医。

以包含不可随意控制投入的 DEA 模型为例，在模型中对不可随意控制投入和可随意控制投入进行不同的处理(Banker & Morey, 1986)。以 VRS 模型为例，其线性规划式为：

$$\begin{aligned}
&\min \theta \\
&s.t.\ X_D\lambda - \theta x_{\mathrm{k},D} \geq 0 \\
&X_N\lambda \leq x_{\mathrm{k},N} \\
&Y\lambda \geq y_k \\
&e\lambda = 1 \\
&\lambda \geq 0
\end{aligned} \tag{6.17}$$

D 表示可随意控制的投入(discretionary inputs)，N 表示不可随意控制的投入（non-discretionary inputs）。

在约束条件中不可随意控制的投入与效率值 θ 不发生直接关系，这使得不可随意控制的投入的作用仅限于保证参比的合理性（投影点的生产环境不优于被评价 DMU），其原理类似于统计分析中对混杂因素的控制。

在文献中，不同的学者对这一类 DEA 模型有不同的命名。最先提出这一模型的 Banker & Morey (1986)将其称为具有外生不可随意控制投入和产出的 DEA 模型(exogenously fixed inputs and outputs beyond the discretionary control of DMU managers)；J. Zhu (2009)将这一模型称为特定测量模型（measure specific model）；William W Cooper et al. (2007)称之为不可随意控制变量模型(nondiscretionary variable model)。

在 MaxDEA 软件中，建立包含不可随意控制的投入/产出的模型只需在选择好 DEA 模型距离函数时同时勾选旁边的“不可随意控制的投入/产出”，然后点击“设置”，即可设置哪些投入产出指标是不可随意控制的变量。

第7章 广义 DEA 模型

从本章开始，内容不再涉及距离函数类型等 DEA 模型本身的选项。本章的广义 DEA 及下一章的面板数据 DEA 讨论的都是 DEA 模型中被评价集和参考集的关系问题。

7.1 广义 DEA 模型

马占新 (2002)提出了样本 DEA 的概念，马占新 & 马生昀 (2009)以样本 DEA 为基础提出了广义 DEA（Generalized DEA）的概念[i]。

广义 DEA 模型的核心是“被评价单元集”与“参考集”的分离。传统 DEA 评价的是决策单元之间的相对有效性，参考集就是被评价单元集自身；而广义 DEA 是对被评价单元与参考集之间的比较，参考集与被评价单元集的关系可能是包含、相等、相交或无关[ii]。

从生产可能集的概念上区分，传统DEA的生产可能集是由被评价集构成的；而广义DEA的生产可能集是由参考集构成的，实现了评价对象与参考标准的分离。例如不同分组的决策单元集与同一固定参考集的比较。

我们用 V 表示被评价集（the eValuated set），用 T 表示参考集（the referred Technology or reference set），在广义 DEA 中被评价集（$k \in V$ ）和参考集（ $j \in T$ ）的关系包括：

1）$V=T$，被评价集与参考集相同，即传统的 DEA；

2）$V \subset T$，被评价集是参考集的一部分；

3）$\mathrm{T} \subset \mathrm{V}$，参考集是被评价集的一部分；

4）$V \bigcap T=\varnothing$，参考集与被评价集互不包含，没有交叉；

5）$V \bigcap T \neq \varnothing$，参考集与被评价集有交叉。

以投入导向 CRS 径向模型为例，广义 DEA 模型可简单地表示为

[i] 在马占新 (2002)提出了的样本 DEA 模型中，参考集被称为样本集。马占新 (2002)提出的“样本 DEA”和马占新 & 马生昀 (2009)提出的“广义 DEA”，其核心都是“决策单元集”与“参考集”的分离，或者说是评价对象与比对标准的分离。从名称上讲“广义 DEA”更为恰当。

[ii] Cook, Seiford, & Zhu (2004)提出的 variable-benchmark DEA 模型的核心也是“决策单元集”与“参考集”的分离，从这个意义上讲 Cook et al. (2004)提出的 variable-benchmark DEA 与马占新 (2002)和马占新 & 马生昀 (2009)提出的样本 DEA 和广义 DEA 是基于同一概念，但马占新 & 马生昀 (2009)对广义 DEA 内容的发展更为丰富。

$$\min \theta$$

$$s.t.\ \sum_{j\in T}\lambda_j x_{ij} \le \theta x_{ik}$$

$$\sum_{j\in T}\lambda_j y_{rj} \ge y_{rk}$$

$$\lambda \ge 0$$

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j\in T, i\in V \qquad (7.1)$$

T 表示参考集，代表参比的技术。

虽然广义 DEA 的提出是基于径向 DEA 模型，但是这一思想可以推广至其它距离函数类型，例如方向距离函数、SBM(ERM)、MinDS、MinDW、成本、收益、利润等模型。

从被评价 DMU 个体的角度考虑，如果被评价 DMU 不在参考集内，即 $k\notin T$，并且也不在参考集 T 的生产可能集范围内，即 $\mathrm{k}=(\mathrm{x_k},\mathrm{y_k}) \notin \{(\mathrm{x,y}): x\ge\sum_{j\in T}\lambda_j x_{ij}, y\le\sum_{j\in T}\lambda_j y_{rj}\}$，则广义 DEA 实质上为超效率模型。

需要注意的是，有些距离函数的标准效率模型的规划式与超效率模型的规划式并不相同，这时广义 DEA 的表达式就需要分别表示。以 SBM(ERM)模型为例，其广义 DEA 模型可表示为

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

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

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

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

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j\in \mathrm{T}, i\in V \qquad (7.2)$$

如果 DMU K 在模型中无可行解，(7.2)，需进一步求解模型

$$\min \rho_{\mathrm{SE}}=\frac{1+\frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{ik}}{1-\frac{1}{s}\sum_{r=1}^{s} s_r^+ / y_{rk}}$$

$$s.t.\ \sum_{j\in \mathrm{T}} x_{ij}\lambda_j - s_i^- \le x_{ik}$$

$$\sum_{j\in \mathrm{T}} y_{rj}\lambda_j + s_r^+ \ge y_{rk}$$

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

$$i=1,2\ldots,m;\ r=1,2\ldots,q;\ j\in \mathrm{T},\ i\in V \qquad (7.3)$$

7.2 群组参比模型



图 7-1 群组参比模型

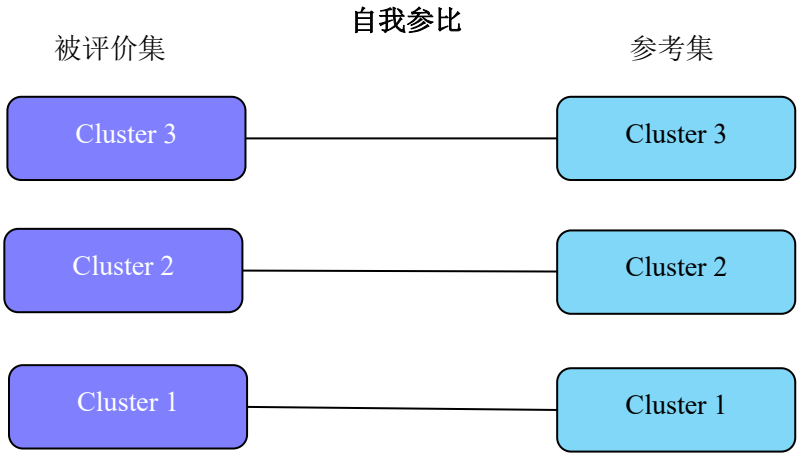
如果数据中的 DMU 按照某种属性可以分为若干群组，例如将中国大陆 31 个省（市、自治区）按东、中、西部划分为三组；或者对现有数据进行某种处理，生成不同群组的 DMU，例如利用广义 DEA 方法中的前沿面平移方法，生成多个新的 DMU 群组。然后在不同的群组之间，按照“被评价集”和“参考集”角色不同，可进行各种形式的参比(Cluster Benchmarking)。

在以下图示中，一条线代表一对被评价集和参考集，右侧的大括号表示对应集合的并集。

7.2.1 自我参比(Self-benchmarking)

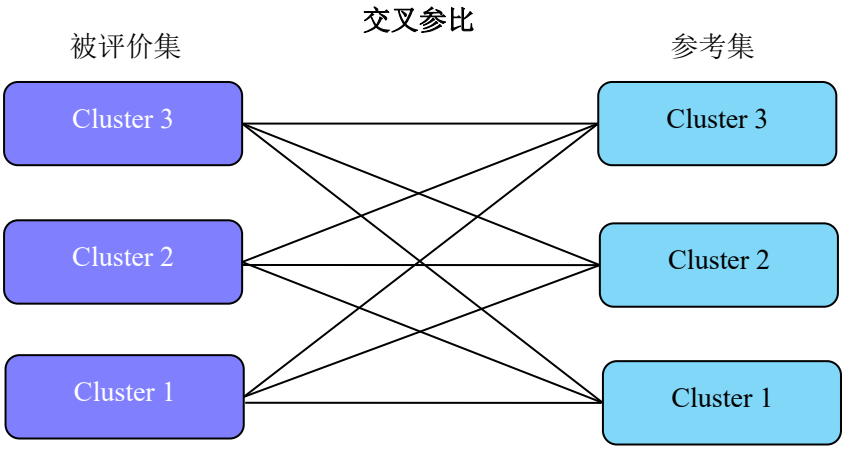
每个群组各自以自身作为“参考集”，分别进行“自评”。如果将各个群组的数据分别保存在不同的数据表中，然后对每个数据表分别进行分析，得到的结果与自我参比的结果相同。

利用自我参比方法，可以方便地进行 MetaFrontier-DEA(共同边界 DEA)分析。通过比较不分组的 DEA 模型的结果（即传统方法，所有 DMU 作为一个参考集）和划分群组后自我参比的结果，即可得到技术缺口比率（technology gap ratio, TGR）。



7.2.2 交叉参比（Cross-benchmarking）

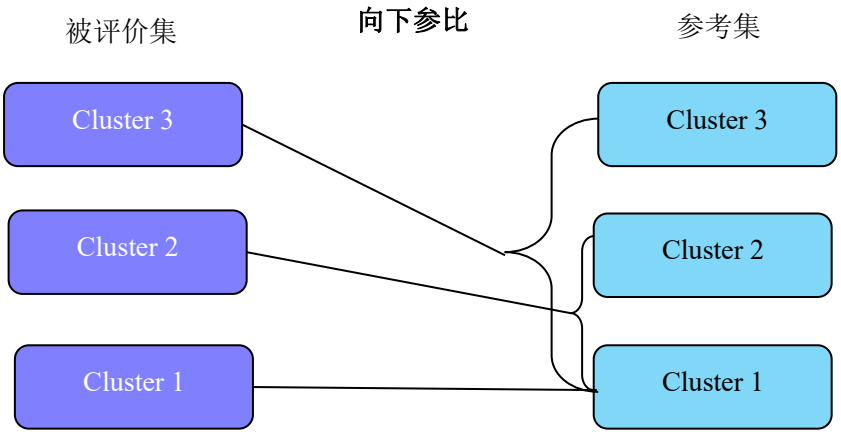
每个群组分别作为参考集，对各个群组依次进行评价。交叉参比包含了自我参比。



7.2.3 向下参比（Downward-benchmarking）

向下参比是每个群组以自身加上群组编号比自己小的所有群组作为参考集。以 Cluster 3 为例，其参考集为三个集合的并集：Cluster 3 ∪ Cluster 2 ∪Cluster 1。群组编号最小的群组的参考集是其自身。

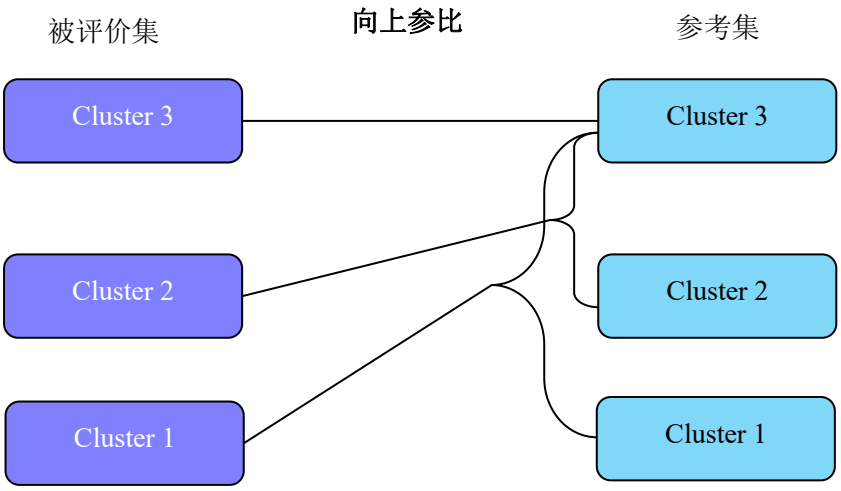
向下参比方法适用于，按照外部因素对生产效率影响的大小将所有 DMU 分为若干组(负面影响越大，群编号越低)，每个群组的参考集是自身加外部影响更大的群组，也就是说参考集的生产环境不会优于被评价集。



7.2.4 向上参比（Upward-benchmarking）

与向下参比相反，向上参比是每个群组以自身加上群组编号比自己大的所有群组作为参考集。以 Cluster 1 为例，其参考集为三个集合的并集：Cluster 3 ∪ Cluster 2 ∪Cluster 1。群组编号最大的群组的参考集是其自身。

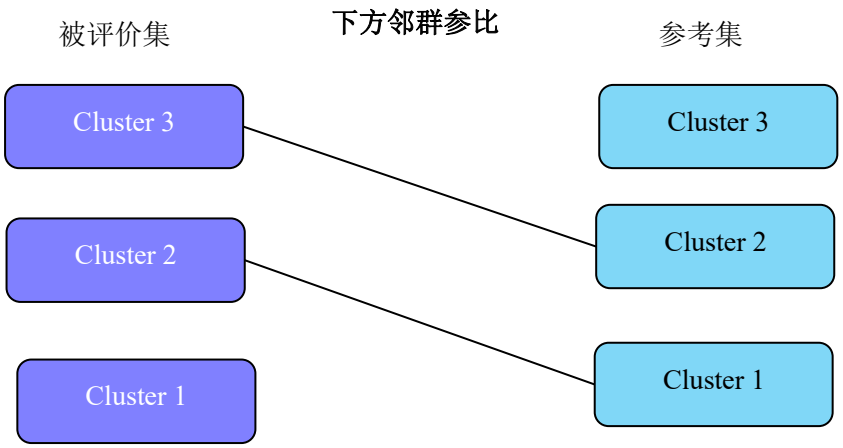
向上参比同样适用于，按照外部因素对生产效率影响的大小将所有 DMU 分为若干组(注意：负面影响越大，群编号越高)，每个群组的参考集是自身加外部影响更大的群组，也就是说参考集的生产环境不会优于被评价集。



7.2.5 下方邻群参比（Lower-adjacent-benchmarking）

与下方的邻群参比是指，每个群组以编号比自己小的相邻**一个**群组作为参考集。例如如果共有 3 个群组，那么 Cluster 3 的参考集是 Cluster 2；Cluster 2 的参考集是 Cluster 1；Cluster 1 不作为被评价群组。

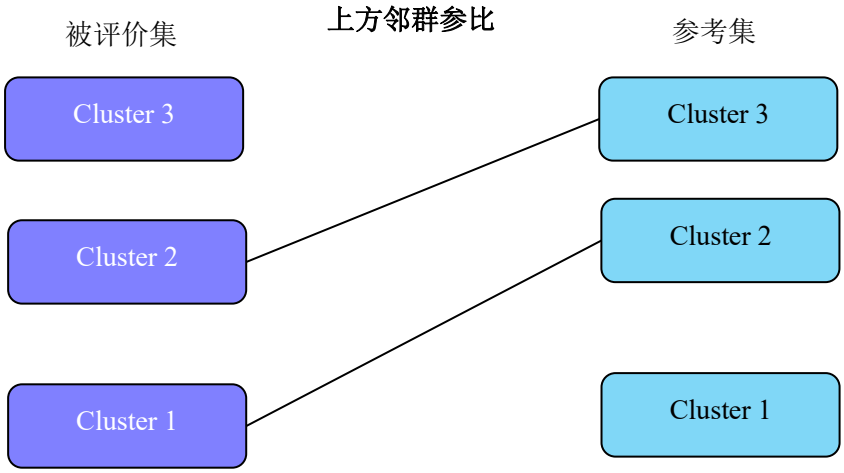
可以通过这一方法，获得面板数据 Malmquist 模型的分析结果中没有提供的分析细节，详见后面对 Malmquist 模型的介绍。



7.2.6 上方邻群参比（Upper-adjacent-benchmarking）

与上方的邻群参比是指，每个群组以编号比自己大的相邻一个群组作为参考集。例如如果共有 3 个群组，那么 Cluster 1 的参考集是 Cluster 2；Cluster 2 的参考集是 Cluster 3；Cluster 3 不作为被评价群组。

这一方法的功能也是可以获得面板数据 Malmquist 模型的分析结果中没有提供的分析细节，详见后面对 Malmquist 模型的介绍。

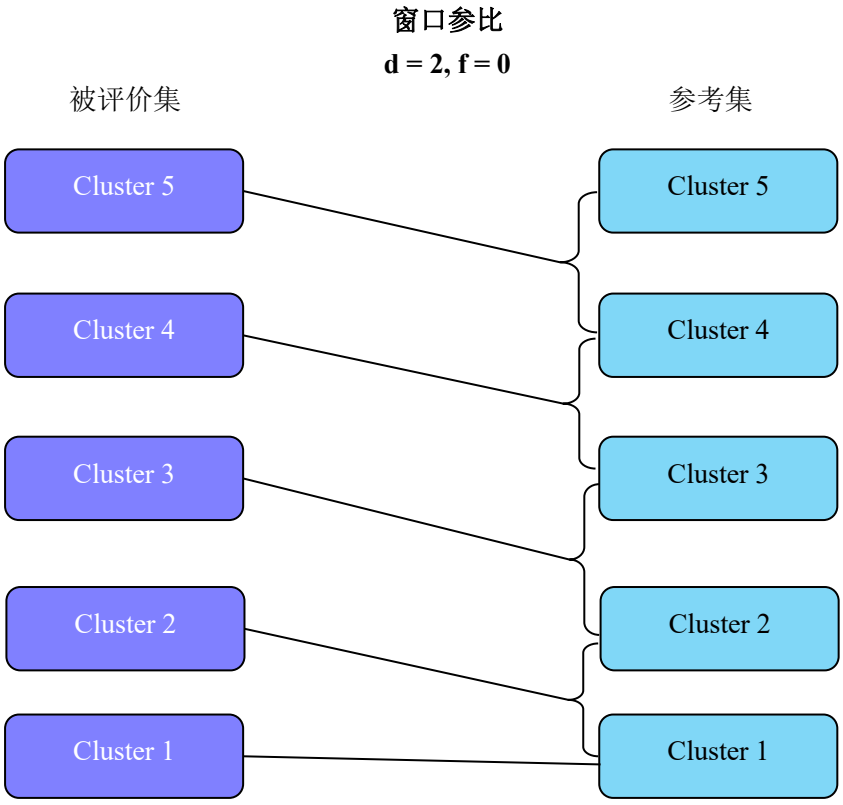


7.2.7 窗口参比(Window-benchmarking)

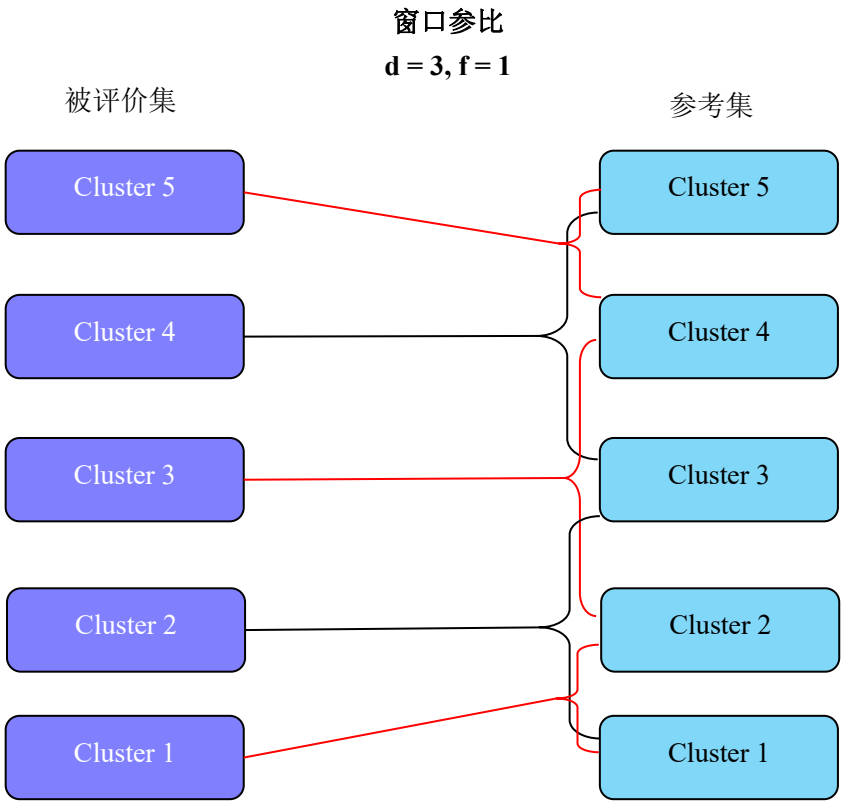
在窗口参比方法中，每个群组以一组编号连续的群组的并集作为参考集，参考集包含被评价群组自身，其中包含的最大群组数量由窗口宽度（wiDth，用 d 表示）决定。窗口默认由被评价 DMU 所在群组及向下的群组构成。可通过设置偏移量来使窗口向上移动，假设被评价 DMU 所在群组的编号为 c，设置偏移量（oFfset，用 f 表示）之后，参考集包括的群组范围为编号 [c−d+1+f, c+f]。

如果共有 5 个群组，窗口宽度 d＝2，偏移量 f＝0，各群组的参考集如下图所示。注意

编号靠近两端（起始和结束）的被评价群组的参考集为不完整的窗口，不完整窗口的数量决定于窗口宽度。



如果窗口宽度为 d=3，偏移量 f=1，各群组的参考集如下：



窗口参比的功能也是可以获得面板数据 Malmquist 模型的分析结果中没有提供的分析细节，详见后面对 Window-Malmquist 模型的介绍。

7.2.8 固定参比(Fixed-benchmarking)

固定参比是指定某一个群组作为参考集。

7.3 群组参比模型的数据格式

对于群组数据，必须包含一个表示群组编号的变量，且变量值必须为整数，例如 1， 2，3……；但可以是不连续的，例如 1， 2， 5， 8……；2001， 2005， 2009……；200101，200102， 200302 等。

在导入群组数据时，在数据定义界面，需要将群组编号变量设为 Cluster。

下面是群组数据的示例：

表 7-1 群组截面数据示例

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

群组数据可以是非平衡的，即每个群组包含的 DMU 数目可以是不同的。对于群组面板数据，每个 DMU 在不同时期必须属于同一群组。例如：DMU A 在时期 1 属于群组 1，则在其它所有时期都必须属于群组 1。

表 7-2 群组面板数据示例

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

第8章 面板数据模型

8.1 面板数据

面板数据（panel data）是指来自两个或更多时期的数据，最常见的面板数据是年度数据，即每个 DMU 在每年都有一组投入产出数据。面板数据可以是连续的年度数据，也可以是不连续时期的数据，例如由两个时期构成的面板数据，第一个时期为 2000 年，第二个时期为 2020 年（表 8-1）。切勿在 DMU 名称后面增加年份，这样做是错误的（表 8-2）。

表 8-1 面板数据格式示例

Period	DMU	Capital	Labor	Product
2000	甲	4323	875	93608
2000	乙	2295	469	225559
2000	丙	6379	1286	327068
2020	甲	6644	1339	201354
2020	乙	1436	297	188926
2020	丙	6281	1266	413738

表 8-2 错误面板数据格式示例

Period	DMU	Capital	Labor	Product
2000	甲 2000	4323	875	93608
2000	乙 2000	2295	469	225559
2000	丙 2000	6379	1286	327068
2020	甲 2020	6644	1339	201354
2020	乙 2020	1436	297	188926
2020	丙 2020	6281	1266	413738

前面介绍的 DEA 模型都是采用一个特定时期的截面数据，分析的是各个 DMU 在该特定时期的技术效率，是针对这一特定时期的生产技术而言的。但是生产是一个长期的连续的过程，在这一过程中，生产技术本身往往是不断发生变化的（技术进步）。

当被评价 DMU 的数据为包含多个时间点观测值的面板数据时，就可以对生产率的变动情况及技术效率和技术进步各自对生产率变动所起的作用进行分析，这就是常用的 Malmquist 全要素生产率（total factor productivity, TFP）指数分析。

我们用下面的简单示例来理解为什么面板数据不能只做技术效率分析。A 和 B 两个 DMU，在时期 1，A(x=4,y=1)的技术效率为 1，B(x=6, y=1)的技术效率为 0.667；在时期 2，A(x=1, y=4)和 B(x=1.5, y=4)的技术效率没有发生变化，A 还是 1，B 还是 0.667。如果只做技术效率分析，从时期 1 到时期 2，A 和 B 的技术效率没有发生任何变化。但实际上，A 和 B 的产出投入比（即生产率）均大幅提高，时期 2 与时期 1 相比，A 和 B 的生产率均提高了 15 倍（2 期是 1 期生产率的 16 倍），这就是技术进步的作用。

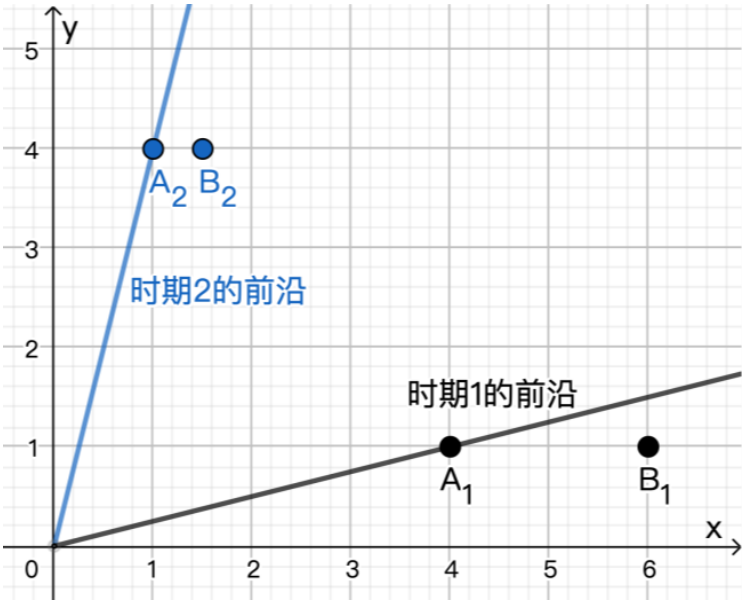


图 8-1 面板数据分析示例

除此之外，窗口（window）DEA 也是常用的面板数据分析方法，它是将一定宽度的时间段（窗口期）内所有 DMU 整体作为参考集，从而使参考集内的数值点（一个数值点可看作是一个“DMU”）的数量成倍增加，窗口 DEA 是解决 DMU 数量不足问题的一种有效途径 (A. Charnes, Clark, Cooper, & Golany, 1985)。

MaxDEA 软件允许使用非平衡面板数据，缺失数据对应的结果也相应缺失。如果是非平衡面板数据，在运行完毕之后会显示警示信息，在日志文件中有更为详细的说明。



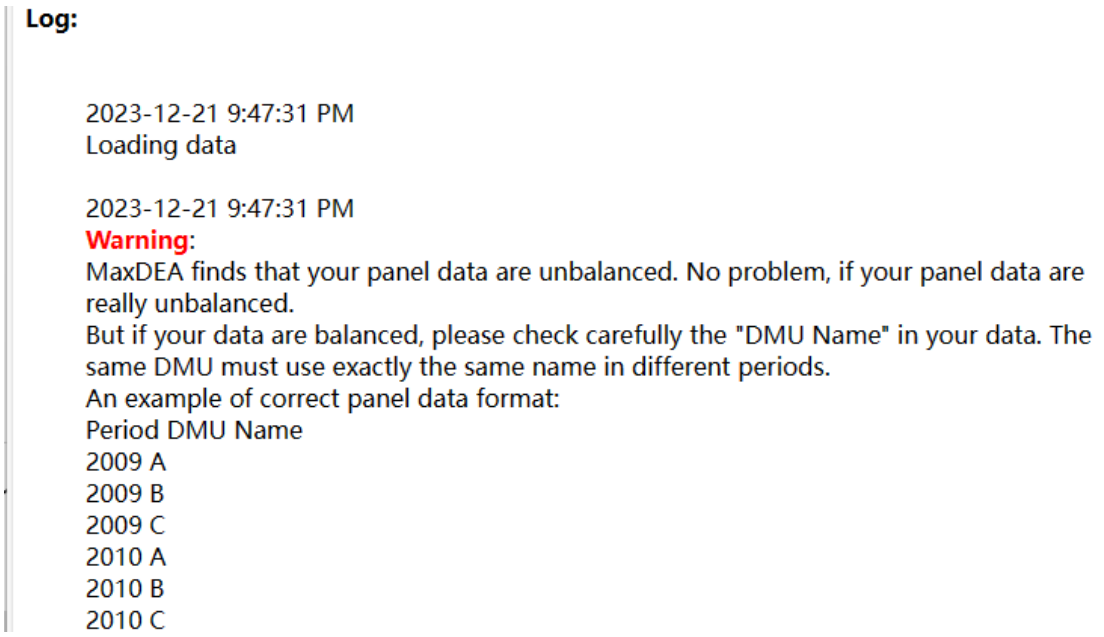


图 8-2 MaxDEA 对使用非平衡面板数据的提示

平衡面板数据是指每个 DMU 在每个时期有且只有一个观测值。如果一个 DMU 在同一个时期出现两个或更多的观测值，例如在表 8-3 中，2020 年有两个 DMU“甲”，则软件会报错（图 8-3）。

表 8-3 错误面板数据示例

Period	DMU	Capital	Labor	Product
2000	甲	4323	875	93608
2000	乙	2295	469	225559
2000	丙	6379	1286	327068
2020	甲	6644	1339	201354
2020	甲	1436	297	188926
2020	丙	6281	1266	413738

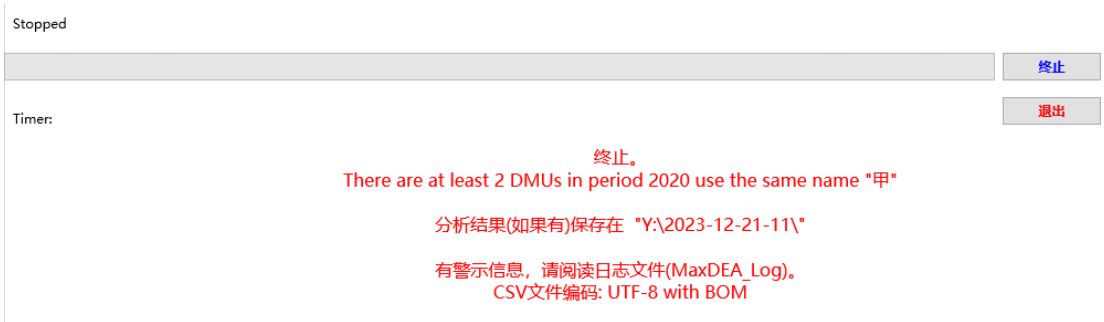


图 8-3 MaxDEA 对错误面板数据的提示

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

图 8-4 在 Excel 中检查面板数据是否平衡

如果你对 MaxDEA 软件给出的非平衡面板的提示有疑问，可以在 Excel 中利用数据透视表功能检查面板数据是否平衡：在 Excel 中插入一个数据透视表，用 DMU 名称变量作为行标签和计数项，用时期变量作为列标签。如果交叉表内的数值均为 1，并且没有空白项，则说明是平衡面板。

如果面板数据特别大，在 Excel 中检查面板数据是否平衡也不容易，可以使用统计软件检查。

Stata 软件中检查面板数据是否平衡(Stata 称为 strongly balanced)的命令：

xtset DMU 名称变量 时期变量

spbalance

xtset 要求两个变量必须是数值型，如果 DMU 名称变量是字符型，则先新建一个 id 变量：

egen id = group(DMU 名称变量)

xtset id 时期变量

spbalance

当数据量较多时，数据平衡与否不是眼睛所能直接做出正确判断的。在整理数据时，千万不要依靠眼睛。

眼见为实吗？看下面这张图，你的眼睛告诉你，A 和 B 两个区域的颜色是不一样的，其实二者颜色完全相同。

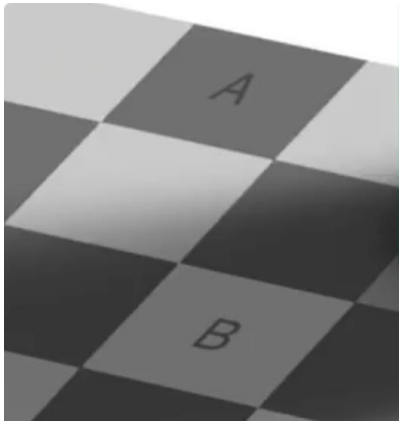


图 8-5 视错觉图片

8.2 生产率变化、技术变化与技术效率变化三者之间的关系

当对面板数据两个时期（t-1,t）进行投入产出分析时，会涉及到三个概念：生产率变化、技术变化和技术效率变化。为了便于理解三者之间的关系，我们先看几种特殊的情形。

(1) 图 8-6：行业技术进步（前沿前移），被评价 DMU K 在两个时期（t-1,t）的生产率不变（投入和产出数值不变）

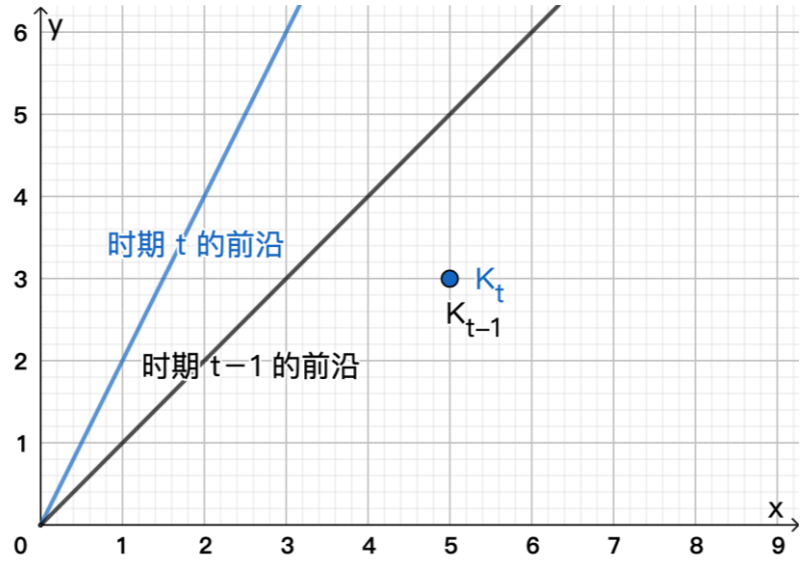


图 8-6 生产率变化、技术变化与技术效率变化之间的关系(1)

生产率是产出与投入的比值，既然 K 的投入和产出均未发生变化，那么 K 的生产率也就不发生变化，K 在两个时期的生产率相同，即生产率指数为 1。由于 t 期的前沿前移，与 t-1 期相比，t 期发生了技术进步，技术变化指数为 2。在 t 期，K 与前沿的距离变得更远，因此其技术效率降低，技术效率指数为 0.5（t 期的技术效率为 0.3，t-1 期的技术效率为 0.6，比值为 0.5）。

需要注意的一点是，技术变化是针对某个时期的行业生产技术而言的，并非是指某个特定 DMU 所采用的生产技术。

在这个例子中，K 在两个时期的生产状况没有发生任何变化，虽然 K 的生产率保持不变，但是由于该生产行业产生了技术进步（前沿生产率提高），从而使 K 的技术效率降低（与前沿的相对距离变大）。

针对多投入多产出的生产过程，虽然 DEA 方法无法计算 K 在两个时期的生产率，但是却可以计算 K 在时期 t 相比时期 t-1 的生产率指数。在实践中，我们往往并不一定需要得到生产率的数值，而是关注生产率的变化情况，即生产率指数。

(2) 图 8-7：行业技术不变，被评价 DMU K 在时期 t 的生产率提高（投入不变但产出增加）

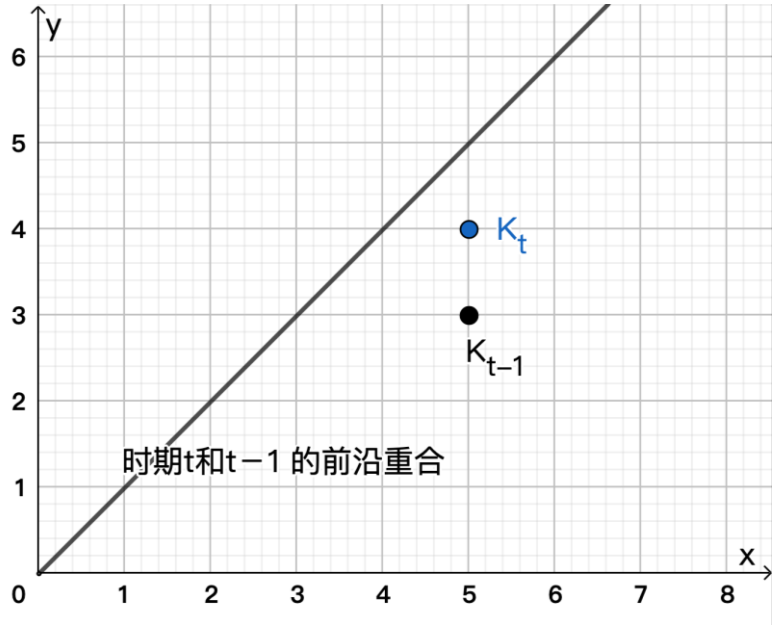


图 8-7 生产率变化、技术变化与技术效率变化之间的关系(2)

K 在时期 t 的生产率提高，生产率指数为 1.333。行业技术没有发生变化（前沿保持不变）。在 t 期，K 与前沿的距离变得更近，因此其技术效率提高，即技术效率指数为 1.333（t 期的技术效率为 0.8，t-1 期的技术效率为 0.6，比值为 1.333）。

(3) 图 8-8：行业技术进步，被评价 DMU K 在时期 t 的生产率提高（投入不变但产出增加）

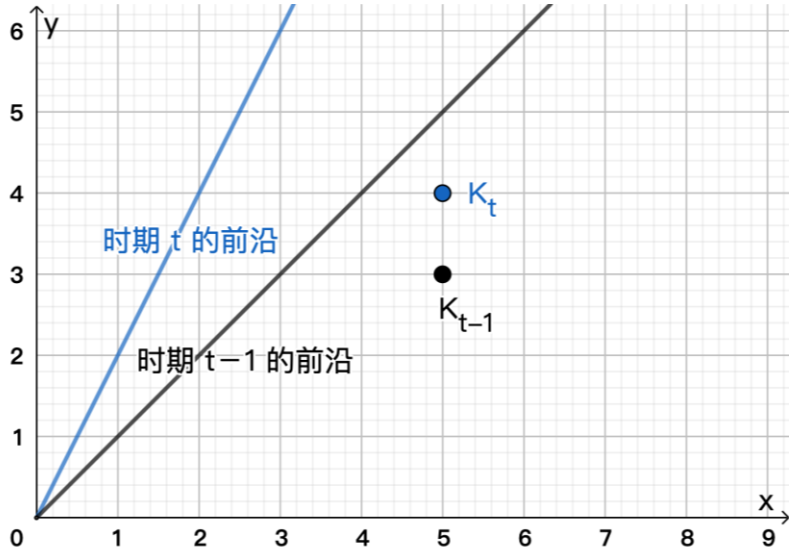


图 8-8 生产率变化、技术变化与技术效率变化之间的关系(3)

K 在时期 t 的生产率提高，生产率指数为 1.333。由于 t 期的前沿前移，即与 t-1 期相比，t 期发生了技术进步，技术变化指数为 2。在 t 期，K 与前沿的距离变得更远，因此其技术效率降低，技术效率指数为 0.667（t 期的技术效率为 0.4，t-1 期的技术效率为 0.6，比值为 0.667）。

8.3 Malmquist 指数的基本原理

面板数据是涉及两个或更多时期的数据，每个 DMU 都有多个时期的投入和产出数值。技术会随着时间而发生变化，这种变化就是通常所说的技术进步，技术进步往往是生产率提高的第一推动力。如果将面板数据拆成单个时期的数据，每个时期分别做分析，可以得出每年的技术效率值。这样做只能得到各个时期的技术效率值，当技术进步是生产率提高的主要推动力的时候，只做技术效率分析是不全面的。

全要素生产率（total factor productivity, 简称 TFP）是指产出数量不能归因于生产要素数量的那部分。从动态的角度更容易去理解 TFP，假设一个生产单位，在生产要素没有发生变化的情况下，与前一年相比其产量增加了 30%。这 30%增加的产量不能归因于生产要素数量，而是由技术进步等生产要素以外的因素所引起的，这就是 TFP 的作用。该生产单位的 TFP 与前一年相比提高了 30%（TFP 指数为 1.3）。

DEA 方法属于非参数方法，无法计算 TFP，但可以计算 TFP 指数。假设一个 DMU 在 t 期和 t-1 期的投入 x 保持不变，其产出分别为 y_t 和 y_{t-1}，则有

$$y_{t-1}=TFP_{t-1}\times f(x) \quad (8.1)$$

$$y_t=TFP_t\times f(x) \quad (8.2)$$

由于生产函数 f(x)未知，无法计算两个时期的 TFP，但其 TFP 指数可以计算，这就是 DEA 方法计算 TFP 指数的基本思路。

$$\text{TFP Index (t-1,t)}=\frac{TFP_t}{TFP_{t-1}}=\frac{y_t}{y_{t-1}} \quad (8.3)$$

Malmquist 全要素生产率指数（Malmquist total factor productivity index，简称 MI）的概念最早源于 Malmquist (1953)，因此将这一类指数命名为 Malmquist 指数。R Färe, Grosskopf, Lindgren, & Roos (1992)最早采用 DEA 的方法计算 Malmquist 指数，并将 Malmquist 指数分解为两个方面的变化：技术效率的变化（technical efficiency change, EC）和生产技术（生产前沿）的变化(technological change, TC)。

不同类型的 Malmquist 指数，其区别仅仅是参比不同的前沿。至于计算 Malmquist 指数采用何种距离类型、何种导向及更复杂的 DEA 模型选项，这属于 DEA 模型的内容，参照前面相关章节。DEA 模型的选择与 Malmquist 指数类型的选择是相互独立的两个问题。为了尽量简化对 Malmquist 指数计算过程的解释，我们采用最简单的 DEA 模型来做图示，即单投入单产出 CRS 模型。

针对被评价 DMU K 的两个时期的投入和产出数据，我们关注的是其在两个时期的 TFP，即 TFP_t 和 TFP_{t-1}。当生产涉及多项投入多项产出时， TFP 无法计算，但是可以计算 TFP

指数，即 MI(t-1, t)。通过 DEA 方法计算 MI (t-1,t)的基本思路为：

$$MI\ (t\text{-}1,\ t)=\frac{TFP(x_t,\ y_t)}{TFP(x_{t\text{-}1},\ y_{t\text{-}1})}=\frac{TFP(x_t,\ y_t)\ /\ TFP(benchmark)}{TFP(x_{t\text{-}1},\ y_{t\text{-}1})\ /\ TFP(benchmark)} \quad (8.4)$$

在前面第 2 章（2.2.2，2.2.4，2.3.2，2.3.4）解释过 DEA 模型得出的效率值的本质是被评价 DMU 的生产率与前沿生产率的比值。因此计算 MI (t-1,t)就转换为计算两个 DEA 效率值的比值，分子是 DMU K 在 t 期的 DEA 效率值，分母是 DMU K 在 t-1 期的 DEA 效率值，但是这两个效率值的计算必须参比同一个前沿。

分子和分母均以 t-1 期的前沿作为参比（图 8-9），$TFP(x_t,\ y_t)\ /\ TFP(benchmark_{t\text{-}1})$为$(x_t,\ y_t)$参比 t-1 期前沿得出的 DEA 效率值，记为 Score_t-1(x_t, y_t)[i]；$TFP(x_{t\text{-}1},\ y_{t\text{-}1})\ /\ TFP(benchmark_{t\text{-}1})$为$(x_{t\text{-}1},\ y_{t\text{-}1})$参比 t-1 期前沿得出的 DEA 效率值，记为 Score_t-1(x_t-1, y_t-1)。

$$MI^{t\text{-}1}(t\text{-}1,t)=\frac{Score_t\text{-}1(x_t,\ y_t)}{Score_t\text{-}1(x_t\text{-}1,\ y_t\text{-}1)} \quad (8.5)$$

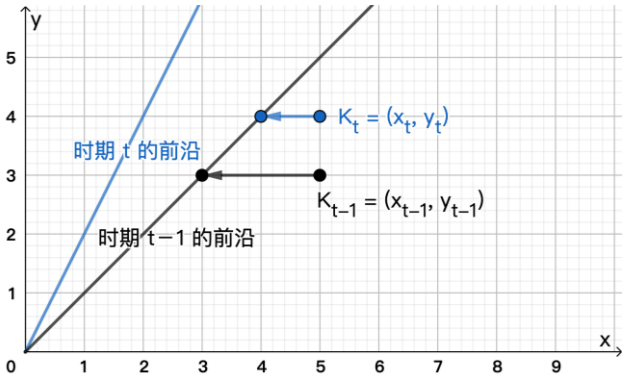


图 8-9 Malmquist 指数示意图(参比 t-1 期的前沿，CRS)

这是参比 t-1 期前沿得出的 MI (t-1,t)。同理，参比 t 期前沿（图 8-10）可以计算得出另一个 MI (t-1,t)。

$$MI^{t}(t\text{-}1,t)=\frac{Score_t(x_t,\ y_t)}{Score_t(x_t\text{-}1,\ y_t\text{-}1)} \quad (8.6)$$

[i] MaxDEA 软件的输出结果为纯文本格式，无法采用上标和下标的方式。为了便于理解软件输出结果，书中的表达式与软件输出结果保持一致。读者自己在撰写论文时，建议参照原始参考文献中所使用的表达式。另外，本书和 MaxDEA 软件中均采用(t, t-1)表示相邻的两个时期，与其他文献中采用(t+1, t)的表达方式相比，只是符号上的差别。

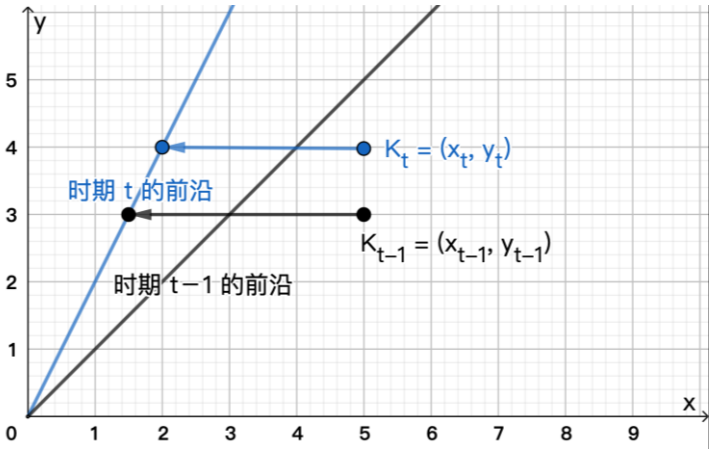


图 8-10 Malmquist 指数示意图(参比 t 期的前沿，CRS)

R Färe et al. (1992)根据 Caves, Christensen, & Diewert (1982)计算 Malmquist 指数的方法，采用上述两个 Malmquist 指数的几何平均值作为被评价 DMU 的 Malmquist 指数，即

$$MI(t\text{-}1,t)=\sqrt{MI^{t\text{-}1}_{(t\text{-}1,t)}\times MI^{t}_{(t\text{-}1,t)}}$$

$$=\sqrt{\frac{Score_t\text{-}1(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t(x_t, y_t)}{Score_t(x_t\text{-}1, y_t\text{-}1)}} \qquad (8.7)$$

$Score_t(x_t, y_t)$ 和 $Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)$ 分别是 K 在 t 期和 t-1 期的技术效率值，其比值是两个时期的技术效率变化（technical efficiency change，简称 EC）

$$EC(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \qquad (8.8)$$

Malmquist 指数可以分解为技术效率变化和技术变化（technological change，简称 TC），即 MI = EC * TC。

$$MI(t,t\text{-}1)=\sqrt{\frac{Score_t\text{-}1(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t(x_t, y_t)}{Score_t(x_t\text{-}1, y_t\text{-}1)}}$$

$$=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\times\sqrt{\frac{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}{Score_t(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t\text{-}1(x_t, y_t)}{Score_t(x_t, y_t)}}$$

$$=EC(t\text{-}1,t)\times TC(t\text{-}1,t) \qquad (8.9)$$

$$TC(t\text{-}1,t)=\sqrt{\frac{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}{Score_t(x_t\text{-}1, y_t\text{-}1)}\times\frac{Score_t\text{-}1(x_t, y_t)}{Score_t(x_t, y_t)}} \qquad (8.10)$$

在 MaxDEA 软件的输出结果中，无论是投入导向还是产出导向，Malmquist 指数的含义

均为，大于 1 表示生产率提高，小于 1 表示生产率降低；效率变化（EC）和技术变化（TC）的含义也是如此。

需要注意的是，在实际应用中，被评价 DMU 往往来自一个较小的样本，即使从理论上讲，被评价生产技术是随着时间不断进步的，但是样本的前沿未必一定前移， DEA 模型得出的技术变化并不一定是技术进步。

8.4 各类 Malmquist 指数计算方法



图 8-11 MaxDEA 软件中 MI 指数类型

MaxDEA 软件提供了 6 类共 9 种 Malmquist 指数，其区别在于参比前沿。Malmquist 指数根据其参比前沿的方法可以分为两大类。

8.4.1 不同时期 MI 的计算参比不同时期的前沿

前面介绍的 Malmquist 指数的计算方法是分别参比相邻两期的前沿（t 期和 t-1 期），得出两个 TFP 指数，然后取其几何平均值作为 Malmquist 指数。例如计算 MI(2019,2020)采用的是 2019 年和 2020 年的前沿，计算 MI(2018,2019)采用的则是 2018 年和 2019 年的前沿。采用这种滚动方式进行参比的 Malmquist 指数有三类：

a) 相邻参比(最常用)

d) 序列参比

e) 窗口参比(相邻)

8.4.1.1 相邻参比（双指数平均）



相邻参比是文献中最常用的经典 MI 计算方法，前面介绍的就是这类 MI 计算方法。

相邻参比 MI(t-1,t)涉及的 DEA 效率值及其含义如下。我们用(x_t, y_t)表示 t 期的投入产出数值，相应地 t-1 期的投入产出数值为(x_t-1, y_t-1)；用 Score_t()表示参比 t 期前沿，相应的参比 t-1 期的前沿表示为 Score_t-1()。

Score_t-1(x_t, y_t)：表示(x_t, y_t)参比 t-1 期前沿得出的 DEA 效率值；

Score_t-1(x_t-1, y_t-1)：表示(x_t-1, y_t-1)参比 t-1 期前沿得出的 DEA 效率值；

Score_t(x_t, y_t)：表示(x_t, y_t)参比 t 期前沿得出的 DEA 效率值；

Score_t(x_t-1, y_t-1)：表示(x_t-1, y_t-1)参比 t 期前沿得出的 DEA 效率值。

表 8-4 Malmquist 模型的部分输出结果(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

假设面板数据包含 2003-2011 年共 9 年的数据。MI(t-1,t)是 t 期 TFP 和 t-1 期 TFP 的比值，数据的起始年份为 2003 年，因此最早的 MI 是 MI(2003, 2004)。

结合 MaxDEA 软件的输出结果来理解 MI(2003, 2004)的计算过程。结果中的列 “t(Period)”是表示当前这一行是计算哪一年 MI，即 MI(t-1,t)中的 t 是指哪一年。当 t=2004 时，MI(t-1,t)= MI(2003,2004)，即表中的第 1 行。

$$\text{MI(t-1,t)}=\sqrt{\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}\times\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t-1, y_t-1})}}=\sqrt{\frac{0.801142}{0.726508}\times\frac{0.854874}{0.775555}}=1.1025$$

$$\text{EC(t-1,t)}=\frac{\text{Score_t}(\text{x_t, y_t})}{\text{Score_t-1}(\text{x_t-1, y_t-1})}=\frac{0.854874}{0.726508}=1.1767$$

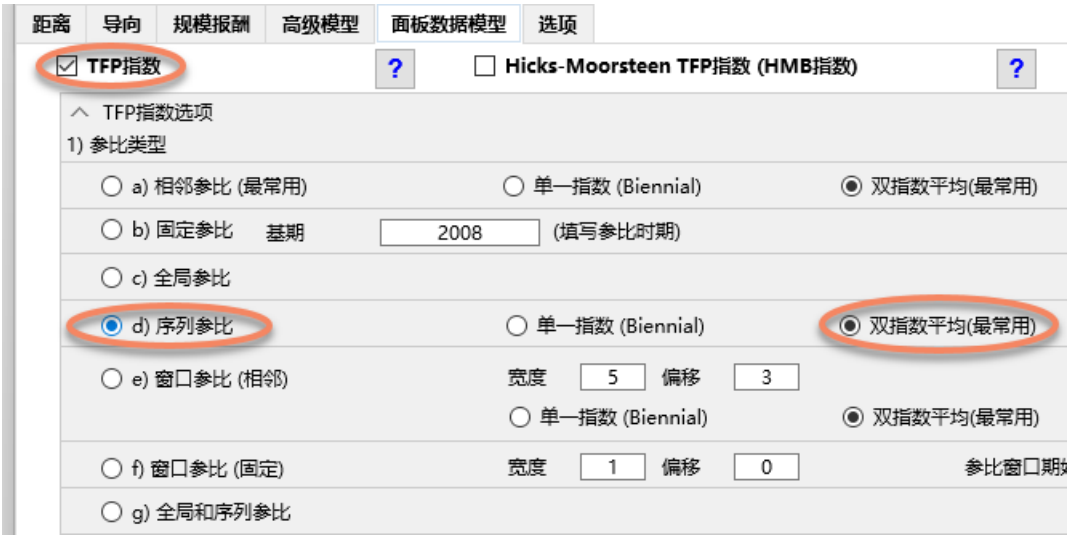
$$\text{TC(t-1,t)}=\sqrt{\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\times\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}}=\sqrt{\frac{0.726508}{0.775555}\times\frac{0.801142}{0.854874}}=0.9370$$

MaxDEA 软件不仅提供了 TFP 指数（MI）及其计算涉及的各种相关 DEA 效率值，并且提供了这些 DEA 模型的投影值等详细结果[i]。

生产率指数（Productivity Index）表中的列 Score_t-1(x_t-1, y_t-1)和列 Score_t(x_t, y_t) 对应的 DEA 模型详细结果均在“表 Score_t(x_t, y_t)”之中。

Score_t-1(x_t-1, y_t-1)，t=2004 和 Score_t(x_t, y_t)，t=2003 表示的是同一个 DEA 效率值；Score_t-1(x_t-1, y_t-1)，t=2005 和 Score_t(x_t, y_t)，t=2004 表示的是同一个 DEA 效率值。

8.4.1.2 序列参比（双指数平均）



序列参比 Malmquist(sequential reference Malmquist)模型是由 Shestalova (2003)提出的一种 MI 计算方法。

序列参比与相邻参比的计算公式完全相同，不再重复。二者的区别在于前沿的构建。相邻参比 t 期的前沿由(x_t,y_t)构建，t-1 期的前沿由(x_t-1,y_t-1)构建，即仅由本期的数据构成；序列参比 t 期的序列前沿由(x_t,y_t), (x_t-1,y_t-1), (x_t-2,y_t-2), …, (x_ 2,y_2), (x_1,y_1)构建，t-1 期的序列前沿由 (x_t-1,y_t-1), (x_t-2,y_t-2), …, (x_ 2,y_2), (x_1,y_1)构建，即序列前沿是由本期及之前的所有时期的数据共同构成。

序列前沿的构建方式决定了 t 期的序列前沿包裹了 t-1 期的序列前沿，因此序列前沿随着时间只会往前推进，不会出现前沿退步的情况，这是序列 Malmquist 模型的重要特征。

当参考集扩大之后，采用径向、SBM(ERM)、方向距离函数、MinDW、混合距离的 DEA 模型，得出的效率值不变或减小。因此，序列 Malmquist 模型得出的技术变化值（TC）不会小于 1，即总是呈现技术进步的结果。

[i] 投影值等详细结果仅在包络模型中提供。

虽然序列前沿不会出现退步，但是如果模型采用 MinDS 距离，计算得出的 TC 却可能会出现技术退步的结果。前面解释过，MinDS 模型的一个特性是当参考集扩大后得出的效率值有可能增大（减小、不变和增加三种情况均存在），这一点是其它类型的距离函数所没有的。所以，当选择 MinDS 距离时，序列 MI 模型得出的技术变化值（TC）可能会小于 1（小于 1、等于 1、大于 1 三种情况都会出现），即仍然会出现技术退步的结果。因此，如果不希望结果中出现技术退步，在序列 MI 模型中不要使用 MinDS 距离类型。

下面我们以三个时期为例，来说明序列前沿的构建。为了方便演示序列前沿的构建原理，在此采用单投入单产出 VRS 模型。图 8-12 是三个时期（1 期、2 期和 3 期）各自的当期前沿。当期前沿由当期的数据构建：1 期前沿为 $A_1B_1C_1D_1E_1$；2 期前沿为 $A_2C_2D_2E_2$；3 期前沿为 $A_3B_3D_3E_3$。

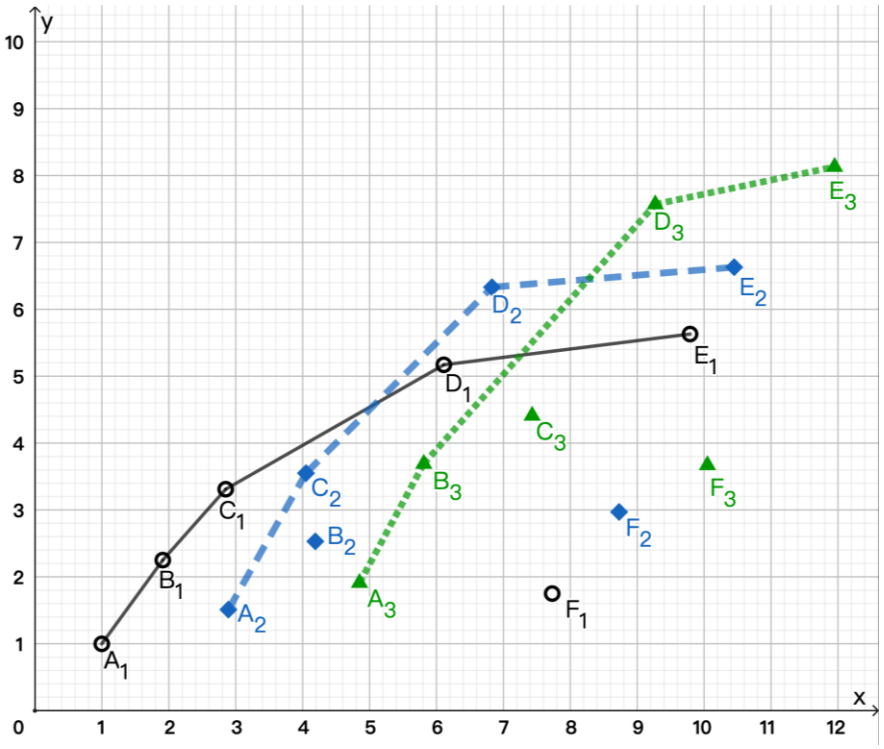


图 8-12 三个时期的当期前沿

时期 1 是第一个时期，其序列前沿就是其当期前沿。图 8-13 是时期 2 的序列前沿 $A_1B_1C_1D_2E_2$，由时期 1 和时期 2 的数据共同构建。

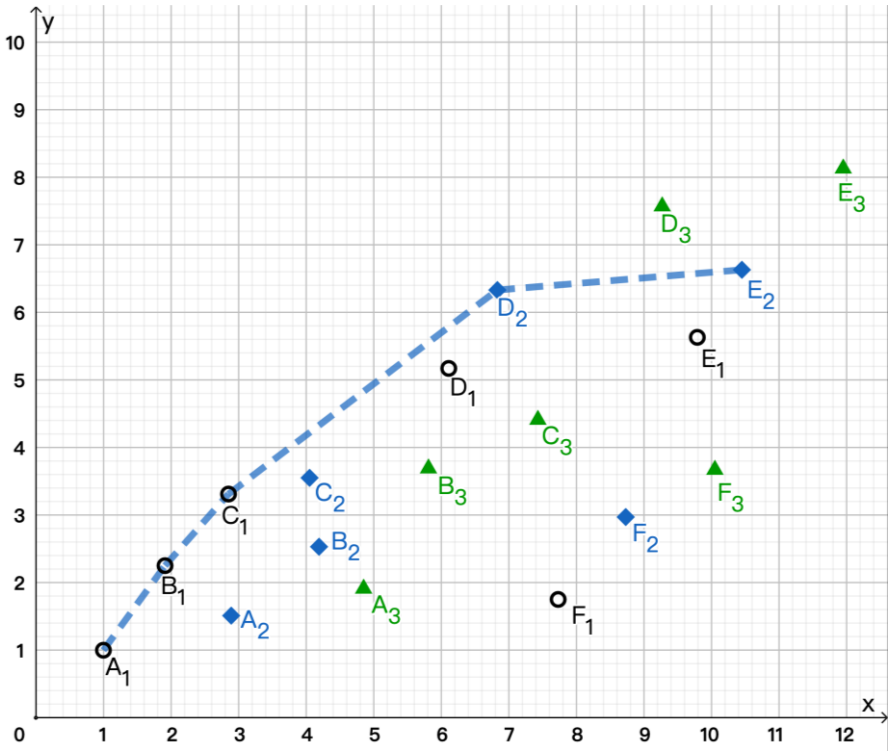


图 8-13 时期 2 的序列前沿

图 8-14 是时期 3 的序列前沿 $A_1B_1C_1D_2D_3E_3$，由时期 1、时期 2 和时期 3 的数据共同构建。

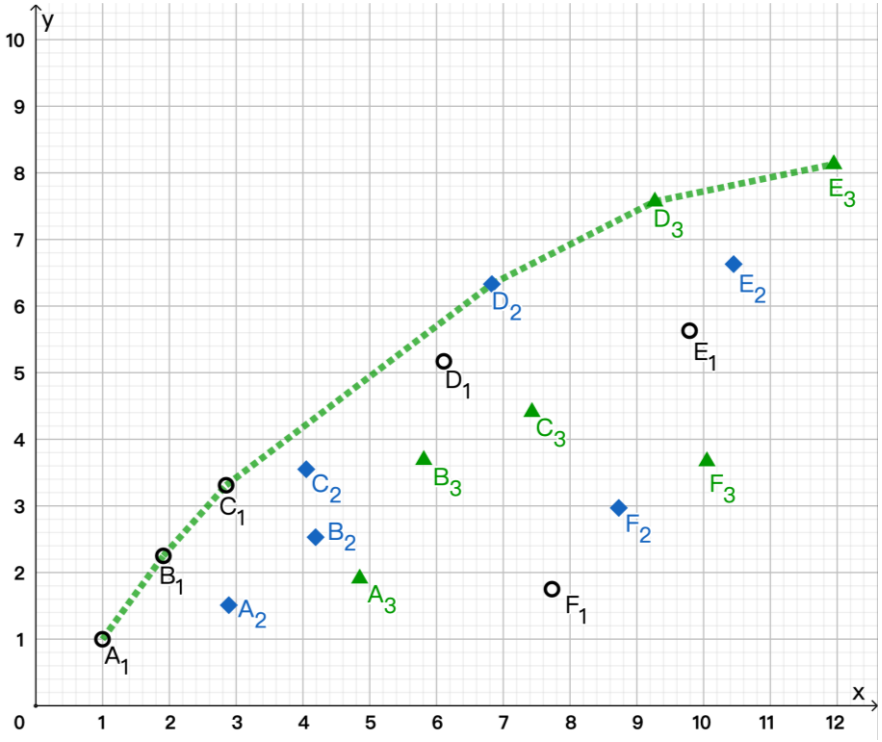


图 8-14 时期 3 的序列前沿

图 8-15 是三个时期的序列前沿的对比图，显示出序列前沿总是进步的特征。

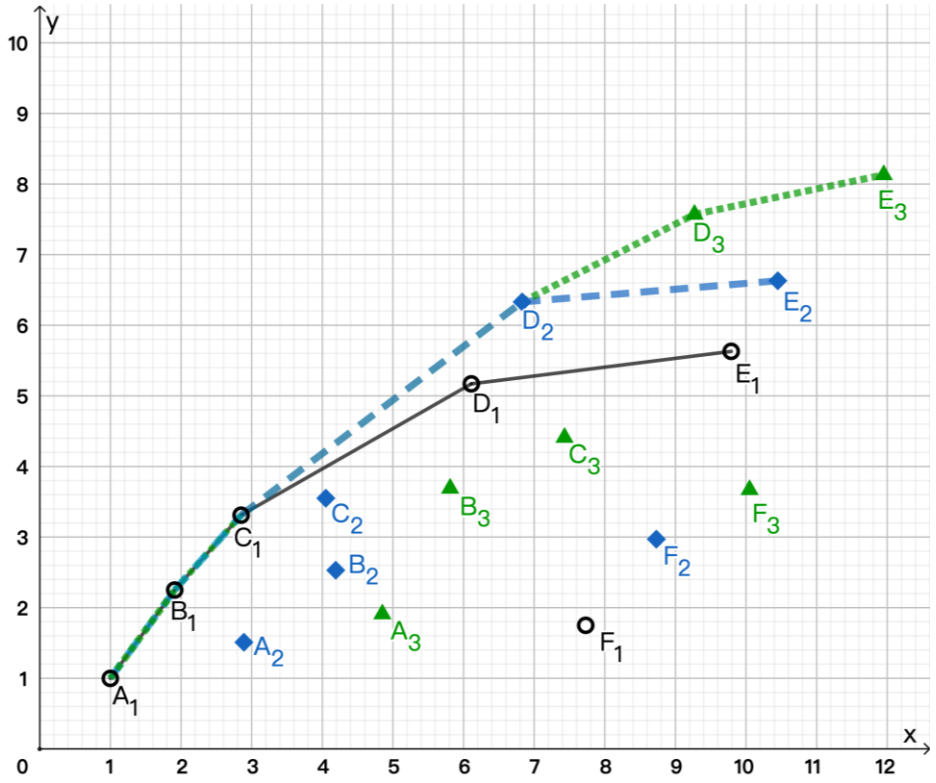
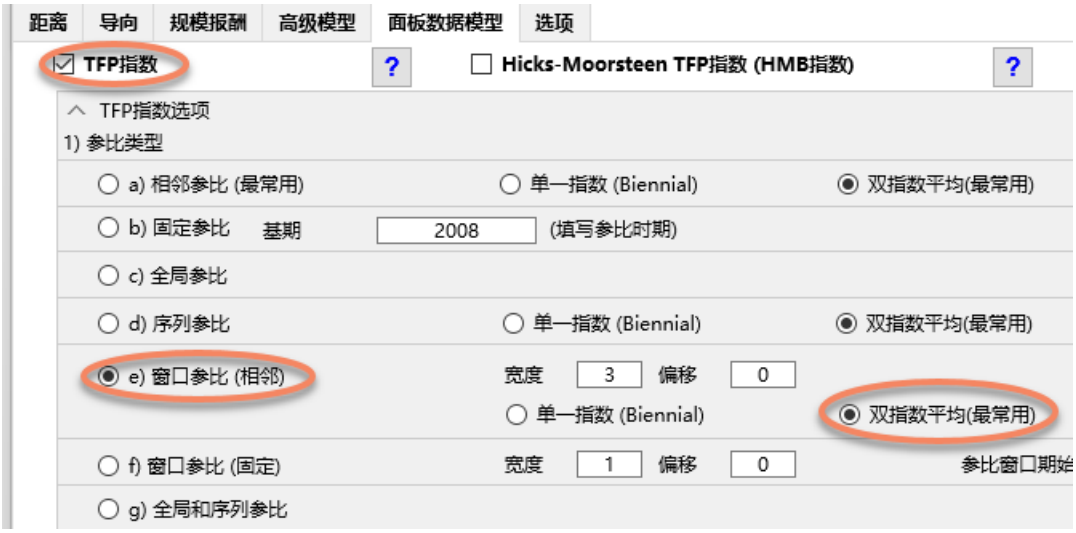


图 8-15 三个时期的序列前沿对比

8.4.1.3 窗口参比（双指数平均）



窗口参比（相邻）与相邻参比的差别也是前沿的构建。t 期的窗口前沿由 t 期及之前的若干时期的数据共同构成，t-1 期的窗口前沿由 t-1 期及之前的若干时期的数据共同构成(Thore, Kozmetsky, & Phillips, 1994; Asmild, Paradi, Aggarwall, & Schaffnit, 2004)。窗口前沿跨越的时期数（包括本期）称为窗口的宽度，例如窗口包括本期及之前的两个时期时窗口宽度为 3。

默认情况下，窗口前沿包含本期及之前的若干时期。如果要包含未来的时期，可以通过窗口偏移来实现。偏移量表示窗口往未来方向偏移的时期数。以窗口宽度=3 为例，2010 年的窗口前沿（默认偏移=0）包含 2010、2009 和 2008；如果偏移=1，则 2010 年的窗口前沿包含 2011、2010 和 2009；如果偏移=2，则 2010 年的窗口前沿包含 2012、2011 和 2010。

以窗口宽度设置为 3 为例，如果不设置窗口偏移（偏移=0），每个时期的窗口包含本期及之前的两期，这样开始的两个时期的窗口宽度就达不到设置的数值 3。第 1 期的实际窗口宽度是 1，只包含本期；第 2 期的实际窗口宽度是 2；第 3 期及之后的各期窗口宽度为 3。如果设置窗口偏移= 1，则每期的窗口包含本期及其前后各 1 期。第 1 期和最后 1 期的实际窗口宽度是 2；其他时期（第 2 期至倒数第 2 期）的窗口宽度为 3。

如果窗口宽度设置为 1，则 t 期的窗口前沿等同于 t 期的单期前沿，窗口参比等同于前面介绍的相邻参比。

各期的窗口未必能够达到设定的宽度。例如偏移=0，窗口宽度设置为 3 时，第 1 期的窗口宽度=1[i]，第 2 期的窗口宽度=2，第 3 期及之后各期的窗口宽度才能达到 3。

如果偏移=0，窗口宽度设置为最大值（总时期数），则 t 期的窗口前沿等同于 t 期的序列前沿，窗口参比等同于序列参比。

在以上介绍的三种 MI 计算方法中，都涉及到跨时期的参比。在计算 MI 所需要的四个 DEA 效率值中，有两个是跨期的参比（t-1 期数据参比 t 期前沿或 t 期数据参比 t-1 期前沿），即 $\text{Score_t}(\text{x_t-1}, \text{y_t-1})$ 和 $\text{Score_t-1}(\text{x_t}, \text{y_t})$。在这两个 DEA 模型中，被评价 DMU 不在参考集中，属于“天然”的超效率模型。

如图 8-16 所示，被评价 DMU $K_t = (x_t, y_t)$参比 t-1 期的前沿时，位于参比前沿的外面，属于超效率模型。

8.4.1.4 对 TC 的进一步分解

对 TC 的进一步分解方法是由 Rolf Färe, Grifell-Tatjé, Grosskopf, & Knox Lovell (1997)提出的。

$$\text{TC(t-1,t)}=\sqrt{\frac{\text{Score_t-1}(\text{x_t-1}, \text{y_t-1})}{\text{Score_t}(\text{x_t-1}, \text{y_t-1})}\times\frac{\text{Score_t-1}(\text{x_t}, \text{y_t})}{\text{Score_t}(\text{x_t}, \text{y_t})}}$$

[i] 偏移=0 时，无论窗口宽度设置为多大，起始期的窗口宽度总是 1。

$$=\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\times\sqrt{\frac{\left(\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\right)}}$$

$$=\text{MATC}(\text{t-1, t})\times\text{BTC}(\text{t-1, t}) \qquad (8.11)$$

MATC(t-1, t) 表示技术变化幅度(magnitude of technological change)，大于 1 表示技术进步(前沿前移)，小于 1 表示技术退步(前沿后退)。使用时期 t-1 的数据(x_t-1, y_t-1)计算得出的 TC，其含义为假定被评价 DMU 投入和产出保持不变，前沿变化引起的生产率的变化。与 TC 相比，MATC 排除了由于在时期 t 投入和产出数值发生变化所引起的技术变化测量发生的偏移。

MATC 与 TC 之间的关系：MATC 可以简单地理解为纯粹的技术变化，而 TC 则包含了混杂因素。

BTC(biased technological change) 反映的是在时期 t 投入和产出数值发生变化所引起的技术变化测量发生的偏移。

BTC 可进一步分解为 OBTC 和 IBTC。

$$BTC=\sqrt{\frac{\left(\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\right)}}$$

$$=\sqrt{\frac{\left(\frac{\text{Score_t-1}(\text{x_t, y_t})}{\text{Score_t}(\text{x_t, y_t})}\right)}{\left(\frac{\text{Score_t-1}(\text{x_t, y_t-1})}{\text{Score_t}(\text{x_t, y_t-1})}\right)}}\times\sqrt{\frac{\left(\frac{\text{Score_t-1}(\text{x_t, y_t-1})}{\text{Score_t}(\text{x_t, y_t-1})}\right)}{\left(\frac{\text{Score_t-1}(\text{x_t-1, y_t-1})}{\text{Score_t}(\text{x_t-1, y_t-1})}\right)}}$$

$$=\text{OBTC(t-1,t)}\times\text{IBTC(t-1,t)} \qquad (8.12)$$

其中，Score_t-1(x_t, y_t-1) 表示参比时期 t-1 的前沿(以时期 t-1 的数据作为参考集)，混合数据虚拟 DMU[x_t, y_{t-1}]的 DEA 效率值。Score_t(x_t, y_t-1) 表示参比时期 t 的前沿(以时期 t 的数据作为参考集)，混合数据虚拟 DMU [x_t, y_{t-1}]的 DEA 效率值。

OBTC(t-1, t) 表示产出偏移技术变化(output biased technological change)，其含义为时期 t 与时期 t-1 相比，产出的变化(即 y_t-1 变为 y_t)使技术变化的测量发生的偏移。

IBTC(t-1, t) 表示投入偏移技术变化(input biased technological change)，其含义为时期 t 与时期 t-1 相比，投入的变化(即 x_t-1 变为 x_t)使技术变化的测量发生的偏移。

8.4.1.5 跨期参比引起的 VRS 模型无可行解问题[i]

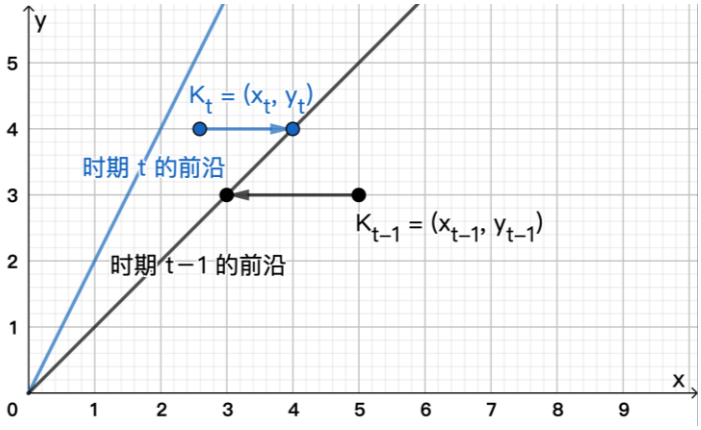


图 8-16 Malmquist 指数计算中的超效率示意图(参比 t-1 期的前沿， CRS)

前面第 5 章介绍过，投入或产出导向的 VRS 超效率模型存在无可行解的问题，同样 Malmquist 模型涉及的 VRS DEA 模型也存在无可行解的问题，即 VRS 模型中 $\text{Score_t}(\text{x_t-1}, \text{y_t-1})$ 和 $\text{Score_t-1}(\text{x_t}, \text{y_t})$ 的计算可能出现无可行解。对无可行解的问题，一种方案是采用第 5 章超效率模型中介绍的解决方法，另一种方案是采用下面介绍的双期前沿参比（单一 TFP 指数）的方法。

8.4.1.6 双期前沿（单一指数）

双期参比（Biennial）就是采用相邻的两期数据(x_t-1,y_t-1)和(x_t,y_t)共同构建的前沿作为参比前沿(Jesús T. Pastor, Asmild, & Lovell, 2011)。因为双期前沿包裹了两个时期所有的被评价 DMU，因此不存在前面所说的被评价 DMU 处于前沿之外的情况，避免了 VRS 模型可能无解的问题。

用 Score_t-1Ut()表示参比双期前沿得出的 DEA 效率值；Score_t-1Ut(x_t, y_t)表示(x_t, y_t)参比双期前沿得出的 DEA 效率值；Score_t-1Ut(x_t-1, y_t-1)表示(x_t-1, y_t-1)参比双期前沿得出的 DEA 效率值。

相邻两期数据参比同一个双期前沿，因此只有一个 TFP 指数，无需再采用计算两个生产率指数然后取其几何平均值的计算方式，双期参比的 MI(t-1,t)为

$$\text{MI(t-1,t)}=\frac{\text{Score_t-1Ut}(\text{x_t}, \text{y_t})}{\text{Score_t-1Ut}(\text{x_t-1}, \text{y_t-1})} \qquad (8.13)$$

EC(t-1,t)的计算与前面相同。虽然计算相邻两期的 MI(t-1,t)参考的是一个双期前沿，但是两期仍然有各自的前沿（单期前沿、序列前沿或窗口前沿），效率变化 EC(t-1,t)的计

[i] DEA 模型无解可能有多种原因，这里特指被评价 DMU 处于前沿以外时 VRS 模型的无解问题。

算仍然是参比当期前沿得出各期的技术效率值，然后取其比值。

$$EC(t\text{-}1,t)=\frac{Score_t(x_t,y_t)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)} \quad (8.14)$$

MI 仍然可以分解为 EC 和 TC。

$$TC(t\text{-}1,t)=\frac{MI(t\text{-}1,t)}{EC(t\text{-}1,t)}$$

$$=\frac{\left(\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}\right)}{\left(\frac{Score_t(x_t,y_t)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)}$$

$$=\frac{\left(\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t(x_t,y_t)}\right)}{\left(\frac{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)} \quad (8.15)$$

其中的 $\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t(x_t,y_t)}$ 反映 t 期前沿与双期前沿之间的距离，比值越大说明 t 期前沿与双期前沿越接近；$\frac{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}$ 反映 t-1 期前沿与双期前沿之间的距离，比值越大说明 t-1 期前沿与双期前沿越接近。二者的比值则反映 t 期前沿与 t-1 期前沿相比的变动情况。

8.4.1.7 双期前沿对 TC 的进一步分解

$$TC(t\text{-}1,t)=\frac{\left(\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t(x_t,y_t)}\right)}{\left(\frac{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)}$$

$$=\left(\frac{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}{Score_t(x_t\text{-}1,y_t\text{-}1)}\right)\times\frac{\left(\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t(x_t,y_t)}\right)}{\left(\frac{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}{Score_t(x_t\text{-}1,y_t\text{-}1)}\right)}$$

$$=MATC\times BTC$$

BTC（biased technological change）反映的是采用间接测量而非直接测量以及在测量 t

期前沿和双期前沿之间的距离时，由于观测点的不同（x_t-1, y_t-1 和 x_t, y_t）而产生的偏移。BTC 也可以参照前面的方法进一步分解为 OBTC 和 IBTC，但其意义不大，没必要再做进一步分解。

当选择序列前沿方法时，因为 t 期的序列前沿包裹了 t-1 期的序列前沿，所以相邻两期的双期序列前沿与 t 期的序列前沿相同，即序列前沿的 Score_t-1Ut() = Score_t()。

双期序列前沿 TC 的计算简化为

$$TC(t\text{-}1,t)=\frac{\left(\frac{Score_t\text{-}1Ut(x_t,y_t)}{Score_t(x_t,y_t)}\right)}{\left(\frac{Score_t\text{-}1Ut(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)}$$

$$=\frac{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}{Score_t(x_t\text{-}1,y_t\text{-}1)}=MATC$$

因此，双期序列前沿的 TC = MATC，BTC = 1，无需进一步分解。**双期序列前沿是唯一不会产生 TC 测量偏移的参比方法**。

8.4.1.8 MI 不可累乘问题

相邻参比的 Malmquist 指数不具备可传递性(circular)，间隔多个时期的 MI(t-n,t) (n>1)不能采用将各时期 MI(t-1,t)相乘的方式来计算，例如

MI(t-2,t) ≠ MI(t-2,t-1) × MI(t-1,t)。

假如面板数据包含 2011 年-2020 年共 10 年的年度数据，则采用前面介绍的 MI 计算方法可以得出 2012 年-2020 年这 9 个年度相比前一年的 MI(t-1,t)。但这 9 年的 MI 相乘，并不等于 2020 年相比 2011 年的 MI(2011,2020)。接下来介绍可以累乘的 MI 计算方法，即所有时期 MI(t-1,t)的计算均参比同一个前沿。

8.4.2 所有时期 MI 的计算均参比同一个前沿

Malmquist 指数的计算采用同一个前沿，这一类型的 Malmquist 指数有三类：

b) 固定参比

c) 全局参比

f) 窗口参比(固定)

固定前沿是指某一固定时期的前沿；全局前沿是指所有时期共同构建的前沿；窗口前沿(固定)是指一个固定窗口期（由多个连续的时期构成的时窗）的前沿。

这三类参比方法得出的 MI、EC 和 TC 具备可传递性，可以累乘，即

MI(t-n,t) = MI(t-n, t-n+1) × MI(t-n+1, t-n+2) × … × MI(t-2,t) × MI(t-1,t)。

8.4.2.1 固定参比（单一指数）



固定参比 Malmquist(fixed reference Malmquist)模型是由 Berg, Forsund, & Jansen (1992)提出的一种 MI 计算方法，它是以某一固定时期的单期前沿作为各期计算 MI(t-1,t)的参比前沿。

所有时期均参比同一固定前沿，因此只有一个 TFP 指数，固定参比的 MI(t-1, t)为

$$MI(t\text{-}1,t)=\frac{Score_f(x_t, y_t)}{Score_f(x_t\text{-}1, y_t\text{-}1)} \quad (8.16)$$

Score_f (x_t, y_t)：表示(x_t, y_t)参比固定前沿得出的 DEA 效率值。

Score_f (x_t-1, y_t-1)：表示(x_t-1, y_t-1)参比固定前沿得出的 DEA 效率值。

EC(t-1,t)的计算方法仍然是：

$$EC(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \quad (8.17)$$

MI 仍然可以分解为 EC 和 TC。

$$TC(t\text{-}1,t)=\frac{MI(t\text{-}1,t)}{EC(t\text{-}1,t)}$$

$$=\frac{\left(\frac{Score_f(x_t,y_t)}{Score_f(x_t\text{-}1,y_t\text{-}1)}\right)}{\left(\frac{Score_t(x_t,y_t)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)}$$

$$=\frac{\left(\frac{Score_f(x_t,y_t)}{Score_t(x_t,y_t)}\right)}{\left(\frac{Score_f(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}\right)} \qquad (8.18)$$

其中的$\frac{Score_f(x_t,y_t)}{Score_t(x_t,y_t)}$反映 t 期前沿与固定前沿之间的距离，比值越大说明 t 期前沿与固定前沿越接近；$\frac{Score_f(x_t\text{-}1,y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}$反映 t-1 期前沿与固定前沿之间的距离，比值越大说明 t-1 期前沿与固定前沿越接近。二者的比值则反映 t 期前沿与 t-1 期前沿相比的变动情况。

8.4.2.2 全局参比（单一指数）



全局参比 Malmquist(global reference Malmquist)模型是由 Jesús T. Pastor & Lovell (2005)提出的一种 Malmquist 指数计算方法，它是以所有各期共同构建的前沿（全局前沿）作为参比前沿。

所有时期均参比同一全局前沿，因此只有一个 TFP 指数，全局参比的 MI(t-1, t)为

$$MI(t\text{-}1,t)=\frac{Score_g(x_t, y_t)}{Score_g(x_t\text{-}1, y_t\text{-}1)} \quad (8.19)$$

Score_g (x_t, y_t)：表示(x_t, y_t)参比全局前沿得出的 DEA 效率值。

Score_g (x_t-1, y_t-1)：表示(x_t-1, y_t-1)参比全局前沿得出的 DEA 效率值。

EC(t-1,t)的计算方法仍然是：

$$EC(t\text{-}1,t)=\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)} \quad (8.20)$$

MI 仍然可以分解为 EC 和 TC。

$$TC(t\text{-}1,t)=\frac{MI(t\text{-}1,t)}{EC(t\text{-}1,t)}$$

$$=\frac{\left(\frac{Score_g(x_t, y_t)}{Score_g(x_t\text{-}1, y_t\text{-}1)}\right)}{\left(\frac{Score_t(x_t, y_t)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\right)}$$

$$=\frac{\left(\frac{Score_g(x_t, y_t)}{Score_t(x_t, y_t)}\right)}{\left(\frac{Score_g(x_t\text{-}1, y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}\right)} \quad (8.21)$$

其中的$\frac{Score_g(x_t, y_t)}{Score_t(x_t, y_t)}$反映 t 期前沿与全局前沿之间的距离，比值越大说明 t 期前沿与全局前沿越接近；$\frac{Score_g(x_t\text{-}1, y_t\text{-}1)}{Score_t\text{-}1(x_t\text{-}1, y_t\text{-}1)}$反映 t-1 期前沿与全局前沿之间的距离，比值越大说明 t-1 期前沿与全局前沿越接近。二者的比值则反映 t 期前沿与 t-1 期前沿相比的变动情况。

由于被评价“DMU”肯定包含在全局参考集内，所以全局参比 Malmquist 指数不存在 VRS 模型无可行解问题。

8.4.2.3 固定窗口参比



“f)窗口参比（固定）”需要设置三个参数，其中窗口宽度和偏移量与“e)窗口参比（相邻）”相同。另一个参数是设置固定窗口的起始时期，终止时期由软件根据窗口宽度自动推定。

固定窗口前沿 MI 的计算与前面固定前沿类似，有两点区别：

1）计算 MI(t-1,t)时，固定前沿是参比某一个固定时期构建的前沿，而固定窗口前沿则是参比一个固定的窗口期构建的前沿。

2）计算 EC(t-1,t)时，固定参比涉及的是相邻两期的单期前沿，而固定窗口前沿则涉及的是相邻两期的窗口前沿。

8.4.2.4 同一前沿对 TC 的进一步分解

全局参比对 TC 的进一步分解：

$$\mathrm{TC(t\text{-}1,t)}=\frac{\left(\frac{\mathrm{Score_g(x_t,y_t)}}{\mathrm{Score_t(x_t,y_t)}}\right)}{\left(\frac{\mathrm{Score_g(x_t\text{-}1,y_t\text{-}1)}}{\mathrm{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}}\right)}$$

$$=\left(\frac{\mathrm{Score_t\text{-}1(x_t\text{-}1,y_t\text{-}1)}}{\mathrm{Score_t(x_t\text{-}1,y_t\text{-}1)}}\right)\times\frac{\left(\frac{\mathrm{Score_g(x_t,y_t)}}{\mathrm{Score_t(x_t,y_t)}}\right)}{\left(\frac{\mathrm{Score_g(x_t\text{-}1,y_t\text{-}1)}}{\mathrm{Score_t(x_t\text{-}1,y_t\text{-}1)}}\right)}$$

$$=MATC(\mathrm{t\text{-}1,t})\times BTC(\mathrm{t\text{-}1,t})$$

固定参比和固定窗口参比对 TC 的进一步分解：

$$\mathrm{TC(t\text{-}1,t)}=\frac{\left(\frac{\mathrm{Score_f\left(x_t,\ y_t\right)}}{\mathrm{Score_t\left(x_t,\ y_t\right)}}\right)}{\left(\frac{\mathrm{Score_f\left(x_t\text{-}1,\ y_t\text{-}1\right)}}{\mathrm{Score_t\text{-}1\left(x_t\text{-}1,\ y_t\text{-}1\right)}}\right)}$$

$$=\left(\frac{\mathrm{Score_t\text{-}1\left(x_t\text{-}1,\ y_t\text{-}1\right)}}{\mathrm{Score_t\left(x_t\text{-}1,\ y_t\text{-}1\right)}}\right)\times\frac{\left(\frac{\mathrm{Score_f\left(x_t,\ y_t\right)}}{\mathrm{Score_t\left(x_t,\ y_t\right)}}\right)}{\left(\frac{\mathrm{Score_f\left(x_t\text{-}1,\ y_t\text{-}1\right)}}{\mathrm{Score_t\left(x_t\text{-}1,\ y_t\text{-}1\right)}}\right)}$$

$$=MATC\mathrm{(t\text{-}1,t)}\times BTC\mathrm{(t\text{-}1,t)}$$

注意 MATC 和 BTC 不具备可传递性，不可累乘。

8.5 各类 MI 比较

1）相邻参比（双指数平均）是应用最早、最多的 MI，虽然存在 VRS 模型无可行解问题，但是可以采用前面介绍的解决 VRS 模型无可行解问题的方法[i]。

2）双期前沿和全局参比 MI 模型不存在 VRS 模型无可行解问题。

3）窗口参比（双指数平均）也存在 VRS 模型无可行解问题。但是当窗口宽度为 w (w≥3)，设置偏移量为 1 至 w-2 之间的数值（例如，当窗口宽度为 3 时，只能设置偏移量为 1；当窗口宽度为 4 时，可以设置偏移量为 1 或 2），则可以避免无解问题。

4）序列参比 MI 模型可以避免出现技术退步的结果（MinDS 距离类型除外）。

5）全局参比、序列参比和窗口参比 MI 的计算相当于增加了“DMU”数量，在 DMU 数量不足时，可采用这三类模型提高结果的区分度。需要注意的是，序列 MI 模型中，“DMU”数量的增加在各期并不均匀，第 1 期没有增加，越往后增加越多，末期增加最多，达到和全局 MI 模型相同的“DMU”数量。在实证研究中，DMU 的数量往往并不是足够多，模型建立的前沿有时候只有几个 DMU 构成，前沿不够精细，得出的结果可靠性或稳定性差。在这种情况下，选择这类可以增加“DMU”数量的 MI 模型，以期达到提高前沿精细度的目的，从而提高结果的可靠性和稳定性。

6）固定参比、全局参比、固定窗口参比三种 MI 可以累乘。t 期的 MI(t-1,t)与之前所有时期的 MI 累乘之后得出的是 MI(0,t)，即 t 期与 0 期（基期或首期）相比的 MI。

$$\prod_{t=1}^{T} MI\left(t-1,t\right)=MI\left(0,\mathrm{T}\right)$$

以全局参比为例，

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）固定参比、全局参比、固定窗口参比三种 MI 模型得出的各期的固定参比效率值 Score_f()或全局参比效率值 Score_g()，参比的是同一个前沿，相互之间数值大小具有可比性。虽然 Score_f()和 Score_g()本身并非 TFP[ii]，但是可以反映 TFP 的相对大小。

MI(t-1,t)是一个 DMU 在 t 期的 TFP 与这个 DMU 在 t-1 期的 TFP 的比值，只能反映这个

[i] 在应用这类 Malmquist 模型的文献中，一般没有提及无可行解问题，实际上往往是在无可行解时将效率值设为 1。

[ii] TFP 数值本身是无法通过 DEA 方法来计算的。

DMU 在两期之间 TFP 的变化率，并不能反映该 DMU 在这两期的 TFP 本身的高低。累乘得出的 MI(0,t) 是一个 DMU 在 t 期的 TFP 与这个 DMU 在 0 期（基期或首期）的 TFP 的比值，只能反映这个 DMU 在 t 期与 0 期（基期或首期）之间 TFP 的变化率，并不能反映该 DMU 在 t 期和 0 期的 TFP 本身的高低。因此，一个 DMU 的 MI(0,t)只能用于自己的 TFP 大小的比较，即该 DMU 在不同时期的 TFP 大小的比较，不能用于不同 DMU 不同时期 TFP 大小的比较。

如果研究者关注的是各 DMU 在各年度的 TFP 增长率，则答案是 MI(t-1,t)；如果还关注各 DMU 的 TFP 累积增加率，则答案是 MI(0,t)。如果研究者要比较不同 DMU 之间不同年度 TFP 的大小，则需要另外的指标。固定参比效率值 Score_f()或全局参比效率值 Score_g()可以用于此目的。

以全局参比为例，全局参比效率值 Score_g(x_t,y_t)的本质是被评价的 DMU 与该 DMU 在全局前沿投影点的 TFP 的比值，其本质也是 TFP 指数，即

$$\text{Score_g}\left(\text{x_t,y_t}\right)=\frac{\text{TFP}(\text{x}_\text{t},\text{y}_\text{t})}{\text{TFP}(\text{benchmark}_g)}=\frac{\text{TFP}(\text{x}_\text{t},\text{y}_\text{t})}{\text{TFP}(\hat{\text{x}}_\text{g},\hat{\text{y}}_\text{g})}。$$

$(\hat{\text{x}}_\text{g},\hat{\text{y}}_\text{g})$ 是被评价 DMU 在全局前沿上的投影点。虽然每个被评价 DMU 在前沿上有不同的投影点，但不同 DMU 在生产前沿上的投影点的 TFP 是等效的。因此，可以认为，不同的 DMU 在计算上述比值时，分母是相同的。或者可以理解为 Score_g(x_t,y_t)是各 DMU 在各时期的 TFP(x_t,y_t)与样本所有年份中的最大 TFP 之间的比值，因此这个比值的最大值是 1。虽然 Score_g(x_t,y_t)并非 TFP，但可以反映不同 DMU 在不同时期 TFP 的相对大小。

利用固定参比效率值 Score_f(x_t,y_t)或全局参比效率值 Score_g(x_t,y_t)可以制作折线图、柱状图等，直观地观察各时期的 TFP 变化情况，例如图 8-17 是 2003 年-2011 年东、中、西部医院 Score_g(x_t,y_t)的对比图(图中三条曲线是东、中、西部区域内各省 Score_g(x_t,y_t)的平均值，因此图中的最大值没有达到 1)。

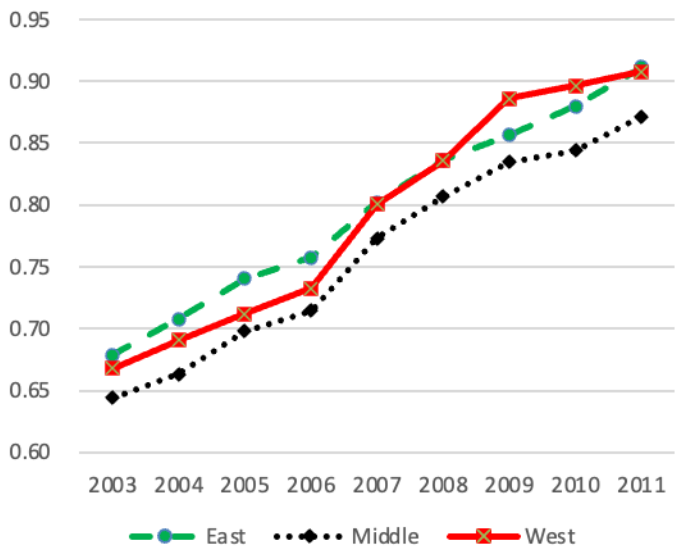


图 8-17 2003 年-2011 年东、中、西部医院的 Score_g(x_t,y_t)

8）如果根据不同 MI 模型分析结果得出的结论差异很大，很可能是由于 DMU 数量不足导致的前沿过于粗糙。从理论上讲，随着 DMU 数量的增加（假设数据没有误差），根据不同类型 MI 分析结果得出的结论会趋于一致。

表 8-5 各类 MI 特点比较

参比方式	指数计算	MI(t-1,t)计算涉及的前沿	各期 MI 是否参比同一前沿	累乘 MI(t-1,t) = MI(0,t)	EC(t-1,t)计算涉及的前沿	TC 进一步分解*	VRS 无可行解问题	可用于不同 DMU 不同时期 TFP 比较的指标
a)相邻参比	双指数平均	单期前沿	否	否	单期前沿	是	存在	无
	单一指数	双期前沿	否	否	单期前沿	否	不存在	无
b)固定参比	单一指数	固定前沿	是	是	单期前沿	否	存在	Score_f(x_t,y_t)
c)全局参比	单一指数	全局前沿	是	是	单期前沿	否	不存在	Score_g(x_t,y_t)
d)序列参比	双指数平均	序列前沿	否	否	序列前沿	是	存在	无
	单一指数	双期序列前沿**	否	否	序列前沿	否	不存在	无
e)窗口参比(相邻)	双指数平均	窗口前沿	否	否	窗口前沿	是	存在，但通过设置窗口偏移可避免***	无
	单一指数	双期窗口前沿	否	否	窗口前沿	否	不存在	无
f)窗口参比(固定)	单一指数	固定窗口前沿	是	是	窗口前沿	否	存在	Score_f(x_t,y_t)

* TC(t-1,t) = OBTC(t-1,t) * IBTC(t-1,t) * MATC(t-1,t)。TC(t-1,t)计算涉及的前沿包括了 MI(t-1,t)计算涉及的前沿和 EC(t-1,t)计算涉及的前沿，不再一一列出。

** t-1 期和 t 期的双期序列前沿等同于 t 期的序列前沿。

*** 当窗口宽度为 3 时，设置偏移量为 1 可避免 VRS 无可行解问题；当窗口宽度为 4 时，设置偏移量为 1 或 2 可避免 VRS 无可行解问题；当窗口宽度为 w 时，设置偏移量为 1 至 w-2 之间的数值可避免 VRS 无可行解问题。DEA 模型无解可能有多种原因，这里特指被评价 DMU 处于前沿以外时 VRS 模型的无解问题。

8.6 文献中常用的 Malmquist 指数分解方法

前面详细介绍了各类 MI 分解为 EC 与 TC 的乘积的计算公式，这种分解并没有考虑规模效率问题。当考虑规模效率时，文献中还有三种常用的 MI 分解方法。在规模报酬界面，选择“6)效率或 TFP 指数分解(包括规模效率/规模效应)”，则在 MaxDEA 软件的输出界面，可以看到四种 MI 分解结果，每种分解结果在一个单独的表中。

FGLR：就是前文详细介绍的 R Färe et al. (1992)的分解方法，不涉及规模效率。

FGNZ：R. Färe, Grosskopf, Norris, & Zhang (1994)的分解方法，在 FGLR 的基础上，对 EC 进一步分解。

RD：Ray & Desli (1997)的分解方法。

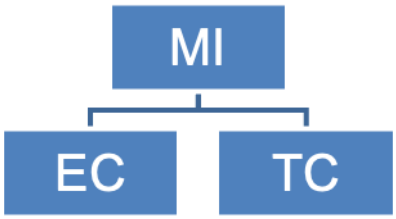
Zofio：Zofio (2007)的分解方法。

如果需要计算每个时期的平均 MI (一个时期所有 DMU 的平均值)或每个 DMU 的平均 MI（一个 DMU 所有时期的平均值），需要使用几何平均。几何平均仍然保持 MI = EC × TC 的分解关系。

注意根据规模效率的定义，计算规模效率需采用径向距离。当采用其他距离类型时，得出的 TE/PTE 比值并不是“规模效率”(Scale Efficiency)，在 MaxDEA 软件中将这一类似比值称为“规模效应”(Scale Effect)，以示区别（详见第 2 章 2.6 规模效率问题）。如果采用的距离类型不是径向距离，则以下分解方法中得出的 SEC 将不是“规模效率变化”，而是“规模效应变化”。

8.6.1 R Färe et al. (1992)的分解方法

R Färe et al. (1992)将 Malmquist 指数分解为技术效率变化（EC）和技术变化（TC）。



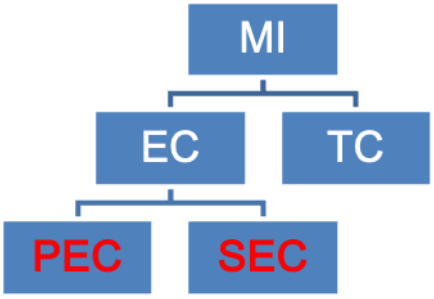
这是最基础的分解方法，这种分解不涉及规模效率。这一分解方法在前面介绍各类 MI 时已经进行详细说明，即

MI(t-1,t) = EC(t-1,t) * TC(t-1,t)

8.6.2 R. Färe et al. (1994)的分解方法

R. Färe et al. (1994)在 R Färe et al. (1992)分解方法的基础上，通过比较 VRS Malmquist 和 CRS Malmquist 模型得出的不同的效率变化值（EC），从而将 R Färe et al. (1992)分解方法中的 EC 进一步分解为纯技术效率变化(Pure Technical Efficiency Change, PEC)和规模效率变化（Scale Efficiency Change, SEC），即

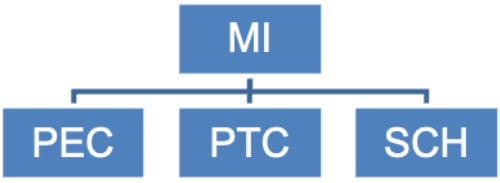
MI = EC * TC = PEC * SEC * TC



8.6.3 Ray & Desli (1997)的分解方法

Ray & Desli (1997)将 CRS 模型得出的 Malmquist 指数分解为纯效率变化（PEC）、纯技术变化（PTC）和规模变化因子（Scale Change Factor），即

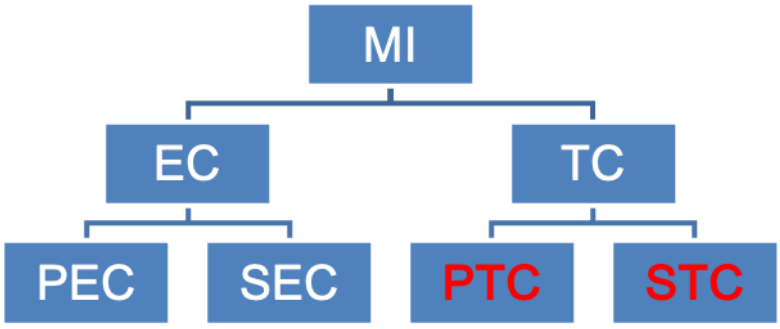
MI = PEC * PTC*SCH



8.6.4 Zofio (2007)的分解方法

Zofio (2007) 的分解方法可以看作是在 R. Färe et al. (1994)分解方法的基础上，将 TC 进一步分解为纯技术变化（PTC）和规模技术变化（STC），即

MI = PEC * SEC * TC＝PEC * SEC * PTC*STC



8.7 Luenberger 指数



在 TFP 指数模型的选项中，还有一项“TFP 指数计算与分解方法”，有两个选项：

1）Malmquist 指数

2）Luenberger 指数

前面介绍的 TFP 指数都是采用比值法计算，称为 Malmquist 指数（MI）。MI 分解为 EC 和 TC 是相乘的关系。MI 大于 1 表示生产率提高，小于 1 表示生产率降低。

如果采用差值法计算 TFP 指数(Chambers, Färe, & Grosskopf, 1996)，则称为 Luenberger 指数（LI），LI 分解为 EC 和 TC 是相加的关系。LI 大于 0 表示生产率提高，小于 0 表示生产率降低[i]。例如相邻参比（双指数平均）Luenberger 指数计算公式为

$$LI(t\text{-}1,t)=\frac{\left(Score_t\text{-}1\left(x_t, y_t\right)-Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)+Score_t\left(x_t, y_t\right)-Score_t\left(x_t\text{-}1, y_t\text{-}1\right)\right)}{2}$$

$$EC(t\text{-}1,t)=Score_t\left(x_t, y_t\right)-Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)$$

$$TC(t\text{-}1,t)=\frac{\left(Score_t\text{-}1\left(x_t\text{-}1, y_t\text{-}1\right)-Score_t\left(x_t\text{-}1, y_t\text{-}1\right)+Score_t\text{-}1\left(x_t, y_t\right)-Score_t\left(x_t, y_t\right)\right)}{2}$$

$$LI(t\text{-}1,t)=EC(t\text{-}1,t)+TC(t\text{-}1,t)$$

Chambers et al. (1996)当时采用差值法是因为该文献中生产率指数的计算采用的是方向距离函数，并用 β^*计算生产率指数，而 β^*既有正数也有负数和 0，不能采用比值的方法计算 TFP 指数。现在这一方法已很少使用。除非必需，不建议使用 Luenberger 指数。

如果需要计算每个时期的平均 LI (一个时期所有 DMU 的平均值)或每个 DMU 的平均 LI（一个

[i] 注意在面板数据模型中如果采用加权加性距离，除 RAM 和 BAM 之外，得出的结果都是“无效率值”，只能采用 Luenberger 指数，并且这种情况得出的 Luenberger 指数意义相反，小于 0 表示生产率提高，大于 0 表示生产率降低。

DMU 所有时期的平均值），需要使用算术平均。算术平均仍然保持 LI = EC + TC 的分解关系。

8.8 Malmquist-Luenberger 指数

Chung et al. (1997)将包含非期望产出的方向距离函数应用于 Malmquist 模型，并将得出的 Malmquist 指数称为 Malmquist-Luenberger 指数(简称 ML 指数)。任何包含非期望产出的 Malmquist 模型得出的 Malmquist 指数都可以称为 Malmquist-Luenberger 指数，也就是说在 MaxDEA 软件中将非期望产出模型与 Malmquist 模型组合，得出的就是 Malmquist-Luenberger 指数。Chung et al. (1997) 采用的是方向距离函数和相邻参比(双指数平均)Malmquist 模型，而在 MaxDEA 软件中，可以选择其它任何距离函数，例如径向、SBM(ERM)、MinDS 等，与 Malmquist 模型进行组合应用；Malmquist 指数也可以选择其它类型，例如全局参比。

注意 ML 指数属于 Malmquist 指数，“TFP 指数计算与分解方法”要选择“Malmquist 指数”。

8.9 Cost-Malmquist 指数

Maniadakis & Thanassoulis (2004)将成本模型与 Malmquist 指数组合应用，得出成本-Malmquist 指数。在 MaxDEA 软件中，还可以将收益模型等与 Malmquist 指数组合应用，得出收益-Malmquist 指数等指标，Malmquist 指数也可以选择其它类型，例如全局参比。

成本 Malmquist 模型的分解，分为两步，MaxDEA 软件直接提供了第一步的分解结果：

第一步：运行成本效率 Malmquist 模型，在结果中有

- Malmquist Index(指成本 Malmquist 指数) = Efficiency Change (指成本效率变化) * Technological Change (指综合技术变化)

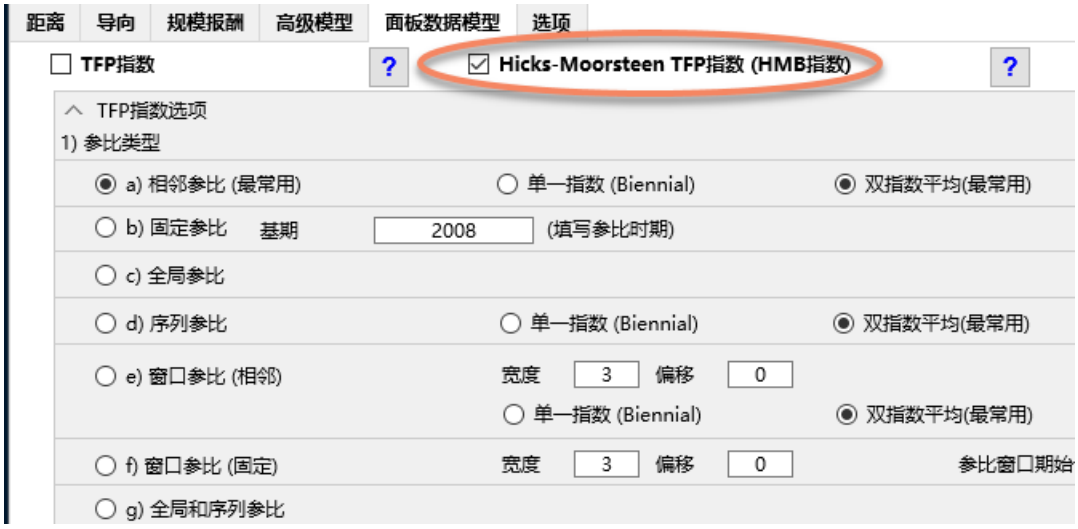
第二步，运行径向 Malmquist 模型，然后利用两步的结果进行以下分解：

- Allocative Efficiency Change（配置效率变化）= Efficiency Change (来自成本 Malmquist 模型) / Efficiency Change (来自径向 Malmquist 模型)
- Price Effect （价格效应）= Technological Change (来自成本 Malmquist 模型) / Technological Change (来自径向 Malmquist 模型)

成本 Malmquist Index 的总分解为

- 成本 Malmquist Index = Efficiency Change（技术效率变化） * Allocative Efficiency Change（配置效率变化）* Technological Change (技术变化) * Price Effect（价格效应）

8.10 Hicks-Moorsteen TFP 指数(HMB 指数)



Hicks-Moorsteen TFP 指数是由 Bjurek (1996)提出的一种 TFP 指数计算方法，因此又被称为 HMB 指数（Hicks-Moorsteen-Bjurek）。

相对于 Malmquist 指数的广泛应用，HMB 指数的应用非常少。这可能有三个方面的原因，一是 HMB 指数的计算更加复杂，Malmquist 指数的计算涉及 4 个组分，而 HMB 指数的计算则涉及 8 个组分，计算复杂度翻倍。二是 HMB 指数的分解更加困难。三是没有专门的软件提供 HMB 指数及其完全的分解。

HMB 指数的一个优点是同时从投入和产出两个角度对 TFP 指数进行测算，不需要设置模型导向。

MaxDEA 9 对促进 HMB 指数的应用做出了两方面的贡献：一是将 HMB 指数由相邻参比推广到固定参比、全局参比、序列参比、窗口参比等所有 Malmquist 指数支持的参比方法；二是给出了与 Malmquist 指数相同的分解方法。从使用者的角度，已经完全可以像使用 Malmquist 指数那样来使用 HMB 指数。

HMB 指数的原理与计算方法在此不予详述，详见软件内的帮助文档。

8.11 对面板数据分析的几点建议

1) 使用含义明确的术语，例如“技术效率”、“纯技术效率”、“规模效率”、“成本效率”、“配置效率”、“全要素生产率”、“全要素生产率指数”等；避免使用含义不明确的术语，例如“效率”、“静态效率”、“动态效率”。

2) 避免使用类似“要素+效率”这样的含义不明确或错误的术语，例如“能源效率”、“能源全要素生产率”、“土地效率”、“土地全要素生产率”等[i]。应该使用诸如“XX 行业全要素生产率”、“XX 行业全要素生产率指数”、“农业全要素生产率”、“农业全要素生产率指数”等含义明确的术语。

3) 单个时期的技术效率 Score_t(x_t,y_t)通常仅用于对 MI(t-1,t)进行分解时计算 EC (t-1,t)。脱离 MI(t-1,t)单独分析 Score_t(x_t,y_t)容易得出误导性的结论。

4) MI(t-1,t)仅用于反映一个 DMU 相邻两期的变化。如果要展示一个 DMU 在多个连续时期的变

[i] 能源和土地都是生产要素，如果单独分析能源和土地的生产率，应该是“能源单要素生产率”和“土地单要素生产率”。

化，建议采用全局参比的累积 MI，即 MI(0,t)；如果要展示多个 DMU 在多个连续时期的变化，建议采用全局参比的 Score_g(x_t,y_t)。

8.12 面板数据的窗口 DEA 模型

窗口 DEA（Window DEA）[i]是将每个 DMU 在各个时期内的数据看作是一个相对独立的“DMU”（在此用双引号表示与实际的 DMU 相区别），假设共有 n 个 DMU，p 个时期，则“DMU”的总数为 np。

如果窗口的宽度为 d（$d \le p$），则窗口的数量为 $w = p - d + 1$。每个窗口内的“DMU”数量为 nd，相当于每个时期的 DMU 数量的 d 倍。增加“DMU”的数量是窗口模型的最明显的功能。

如果窗口宽度设置为 1，窗口内的“DMU”数量最少，等于实际的 DMU 数量，相当于将各时期的数据分离，然后分别对各期的“DMU”进行分析。得出的效率值 Score 等于前面介绍的相邻参比（双指数平均）MI 模型中的 Score_t(x_t,y_t)。

如果窗口宽度设置为面板数据包含的时期数量 p，窗口内的“DMU”数量最多，相当于所有时期的“DMU”汇总成一个总的参考集，然后所有“DMU”均以此为参考集进行分析。得出的效率值 Score 等于前面介绍的全局参比 MI 模型中的 Score_g(x_t,y_t)。

8.13 在不能获得 TFP 的情况下，如何分析 TFP 的影响因素

通过 DEA 方法只能计算 TFP 指数（TFP 变化），不能计算 TFP 本身的数值。如果分析 TFP 变化率的影响因素，可以直接使用 TFP 指数作为回归模型的因变量；如果要分析 TFP 本身的影响因素，其实也是可以使用 TFP 指数作为回归模型的因变量的，但注意需要使用特定的 TFP 指数（或特定的效率值）作为因变量并且采用面板数据固定效应回归模型。下面来介绍如何使用“TFP 指数”作为因变量来分析“TFP”的影响因素。

方法一： 以“累积全局 Malmquist 指数的对数值”作为因变量，采用固定效应回归模型，与以“TFP 的对数值”作为因变量得出的回归系数相同。

方法二： 以“全局效率值的对数值”作为因变量，采用固定效应回归模型，与以“TFP 的对数值”作为因变量得出的回归系数相同。

以上结论读者可以自行推导证明。

具体计算方法如下：

方法一： 以“累积全局 Malmquist 指数的对数值”代替“TFP 的对数值”作为因变量 Y 做回归。

1) 在 MaxDEA 软件面板数据模型中选择“TFP 指数”；“参比类型”选择“c)全局参比”；“TFP 指数计算与分解方法”选择“Malmquist 指数”；“计算哪两个时期的 TFP 指数”选择“(t0, t)”。

2) 对“累积全局 Malmquist 指数”，即 MaxDEA 9.0 输出结果中的 MI(t0, t)，取对数值作为因变

[i] 此处的 Window 在中文 DEA 文献中多翻译为“视窗”，这移植自“视窗操作系统 Windows”。 Windows 操作系统是可视化窗口操作，所以称为“视窗”很贴切。但 DEA 模型中的 Window 是指包含多个时期的一个窗口期，并没有可视的含义。所以翻译为“窗口”或“窗口期”更为恰当。这类模型中文可以称为“窗口数据包络分析（窗口 DEA）”或“窗口期数据包络分析”、 “数据包络分析窗口模型”等。

量 Y。

3) 采用固定效应(fixed-effects)模型做回归，Stata 软件的命令如下：

设置面板数据：xtset　id 变量　时期变量

固定效应模型（可在此基础上增加模型选项）：xtreg　因变量　自变量,　fe

方法二:　以“全局效率值的对数值”代替“TFP 的对数值”作为因变量 Y 做回归。

1) 在 MaxDEA 软件面板数据模型中选择“TFP 指数”；“参比类型”选择“c)全局参比”；“TFP 指数计算与分解方法”选择“Malmquist 指数”或“Luenberger 指数”，二者均可，因为本方法的回归模型并不使用 TFP 指数。

2) 以所有时期（包含 0 期）的全局效率值，即表“Score_g(x_t,y_t)”中的 Score，取对数值作为因变量 Y。

3) 采用固定效应(fixed-effects)模型做回归。

如果要验证方法二的回归系数与方法一的回归系数是相同的，数据不要包含 0 期（首期）的 Score_g，否则方法二与方法一的样本量就会不相等，结果就会不同。但在实际应用中，方法二可以包含 0 期的 Score_g。方法一的因变量使用 TFP 指数，只能是从 1 至 t 期；方法二的因变量使用全局效率值，是从 0 到 t 期，方法二比方法一可以多利用一个时期的数据，因此使用“全局效率值的对数值”代替“TFP 的对数值”作为因变量是更好的方法。

以上介绍的方法选择的参比类型是“全局参比”，也可以选择“固定参比”或“窗口参比（固定）”，在后两种参比类型的输出结果中，与表“Score_g(x_t,y_t)”对应的是表“Score_f(x_t,y_t)”。

可以分别采用“全局参比”、“固定参比”和“窗口参比（固定）”三种方法构建 3 个回归模型，然后比较 3 个模型的拟合效果，最后取其中拟合效果最好的一个模型。

第 9 章 新型两阶段网络 DEA 与广义网络 DEA

数据包络分析(DEA)作为评估决策单元(DMU)相对效率的非参数核心工具，其传统模型是将 DMU 视为内部结构不可见的“黑箱”，仅关注最终投入与产出的转化关系。然而，真实生产系统内部往往包含复杂的子流程、多阶段结构以及相互流转的中间流(Intermediates)。网络 DEA(Network DEA)的提出，正是为了打开这一“黑箱”，分析系统内部各节点(阶段)[i]的运行效率及其对系统整体的影响。在网络 DEA 的技术逻辑中，各个生产阶段在技术上具备明确的独立性，即每个节点各自构建其生产可能集(PPS)。这种技术独立性使得将一个 DMU 在阶段 1 的技术与另一个 DMU 在阶段 2 的技术进行线性组合在技术上是可行且合规的。

尽管两阶段及广义网络 DEA(涵盖多节点、串并联及任意拓扑结构)极大地扩展了传统 DEA 的应用范围，但在理论与算法层面长期面临一个痛点：**系统效率分解的不唯一性，并且分解得出的节点效率值往往并非该节点在其独立 PPS 下的效率值**。在传统网络 DEA 的乘数模型求解过程中，由于可能存在多重最优解，系统总效率与各子节点效率之间的代数映射关系并不稳定。这种不稳定的分解为网络 DEA 模型节点效率值的进一步分析（例如节点的 TFP 指数分析）设置了障碍。为解决这一问题，既往研究多依赖人为设定二次目标、引入主观偏好权重或构建主从博弈模型来强行“锁定”分解结果。

MaxDEA 团队开发的新型两阶段网络 DEA 模型从根本上解决了这一难题，其系统效率与两个阶段的效率之间天然具备唯一性等式关系：**系统整体效率值等于两阶段唯一效率值之乘积。**在此基础上，将其推广到广义网络 DEA 和基于全维有效面(FDEF)的广义网络 DEA。

若您在学术研究或论文中引用本文，请参考以下引文格式：

英文格式： MaxDEA Team. (2026-10-26). *Innovative Two-Stage and Generalized Network DEA*. MaxDEA. https://maxdea.com/2026/202609.pdf.

中文格式： 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]为引用日期。

[i] 本文在两阶段网络 DEA 中，使用“阶段(Stage)”，在广义网络 DEA 中，使用“节点(Node)”，来表示网络的内部结构。另外，对于权重向量 v 和 u，会混用“权重”和“（影子）价格”。

符号及数据说明

公式采用矩阵和向量表达式。为方便读者复制公式，尽量采用**纯文本格式**。

矩阵：X、Y、Z 是投入、产出和中间流矩阵。在乘数模型中矩阵行数等于样本 DMU 数量，在包络模型中矩阵列数等于样本 DMU 数量。

X、Y、Z > 0。

列向量：x_o、y_o、z_o表示当前被评价 DMU 的投入、产出和中间流向量(使用下标 $_o$ 在视觉上更为醒目)。λ 为包络模型中的 DMU 线性组合系数。

行向量：v、u、w 是乘数模型中投入、产出和中间流的权重。

v、u、w、λ，默认为非负变量 (≥ 0)。

用粗体 **1** 表示所有元素均为 1 的行向量或列向量。

$u_o \in R$ 是标量变量，用于 VRS 乘数模型，表示规模报酬常数项(截距项)。

上标：$^{(1)}$ $^{(2)}$ $^{(t)}$ 表示阶段 1、阶段 2、节点 t；$^{(in)}$ $^{(out)}$ 表示中间流在节点 t 中的角色是投入还是产出。

9.1 两阶段网络 DEA (规模报酬不变 CRS)

9.1.1 两阶段网络 DEA 模型

Kao & Hwang (2008)和 Liang, Cook, & Zhu (2008)提出了将系统整体效率定义为两个节点效率乘积形式的两阶段网络 DEA 模型，开创了两阶段网络 DEA 模型的乘积效率分解框架(Multiplicative Decomposition)。

阶段 1 的效率

$E^{(1)}* = \max (w \cdot z_o / v \cdot x_o)$

$s.t. \quad (w \cdot Z) / (v \cdot X) \le 1$ (9.22)

阶段 2 的效率

$E^{(2)}* = \max (u \cdot y_o / w \cdot z_o)$

$s.t. \ (u \cdot Y) / (w \cdot Z) \le 1$ (9.23)

系统整体效率

$E* = E^{(1)} \times E^{(2)} = \max (u \cdot y_o / w \cdot z_o) \times (w \cdot z_o / v \cdot x_o) = \max (u \cdot y_o / v \cdot x_o)$

$s.t. \quad (w \cdot Z) / (v \cdot X) \le 1$

$(u \cdot Y) / (w \cdot Z) \le 1$ (9.24)

利用 Charnes-Cooper 变换，得到如下线性规划(LP)模型（投入导向）：

$\max u \cdot y_o$

$s.t. \ v \cdot X - w \cdot Z \ge 0$

$w \cdot Z - u \cdot Y \ge 0$

$v \cdot x_o = 1$ (9.25)

对偶包络模型

$\min \theta$

$s.t. \ X \cdot \lambda^{(1)} \le x_o \cdot \theta$

$Y \cdot \lambda^{(2)} \ge y_o$

$Z \cdot \lambda^{(1)} \ge Z \cdot \lambda^{(2)}$

$\theta \in R$ (9.26)

两阶段网络 DEA 的系统整体效率可分解为两部分的乘积：

$E^* = (u \cdot y_o) / (v \cdot x_o) = (u \cdot y_o) / (w \cdot z_o) \cdot (w \cdot z_o) / (v \cdot x_o) = E^{(1)} \times E^{(2)} \le E^{(1)*} \times E^{(2)*}$

上述两阶段网络 DEA 框架通过打开系统“黑箱”，成功刻画了系统内部中间产品的流转过程。在其理论架构下，模型表现出以下特征和缺陷：

模型特征 1：中间流价格相等的刚性协同(Rigid Synergy)

乘数模型严格要求阶段 1 中间流（作为产出）的权重 $w^{(out)}$与阶段 2 中间流（作为投入）的权重 $w^{(in)}$必须“一对一”绝对相等。

模型特征 2：技术改进路径的混杂性

刚性协同机制映射到对偶包络模型中，表现为径向与非径向的混合——模型仅对初始投入 x（在产出导向模型中是最终产出 y）施加径向等比例缩放约束(θ)，而中间流 z 则具有完全的自由度。这种混杂结构削弱了模型的经济学含义，并且导致投影偏离了经典的技术改进路径。

缺陷 1：系统效率分解的不唯一性

在此框架下，尽管系统整体效率唯一确定，但由于乘数模型往往存在多重最优解，得出的阶段效率分解值并不具备唯一性。

$E^* = E^{(1)} \times E^{(2)}$，$E^{(1)}$和 $E^{(2)}$不具备唯一性。

缺陷 2：分解得出的阶段效率值可能偏离各自的 PPS

在刚性协同机制下分解得出的阶段效率，不一定是各阶段在各自 PPS 下的真实技术效率。

$E^{(1)} \le E^{(1)*}$，$E^{(2)} \le E^{(2)*}$。

9.1.2 基于中间流虚拟价值守恒的两阶段网络 DEA 模型

针对上述框架存在的痛点，提出如下中间流总价值相等的柔性协同(Flexible Synergy)的两阶段网络 DEA 模型：

$$E^* = \max (u \cdot y_o) / (v \cdot x_o)$$

$$\text{s.t.} \quad (w^{(1)} \cdot Z) / (v \cdot X) \le 1$$

$$(u \cdot Y) / (w^{(2)} \cdot Z) \le 1$$

$$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o} \qquad (9.27)$$

利用 Charnes-Cooper 变换，得到如下线性规划(LP)模型（投入导向）：

$$E^* = \max u \cdot y_o$$

$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z \ge 0$$

$$w^{(2)} \cdot Z - u \cdot Y \ge 0$$

$$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o}$$

$v \cdot x_o = 1$ (9.28)

对偶包络模型

$E^* = \min \theta$

s.t. $X \cdot \lambda^{(1)} \le x_o \cdot \theta$

$\mathbf{Z \cdot \lambda^{(1)} \ge z_o \cdot \tau}$

$\mathbf{Z \cdot \lambda^{(2)} \le z_o \cdot \tau}$

$Y \cdot \lambda^{(2)} \ge y_o$

$\theta, \tau \in R$ (9.29)

或者利用 Charnes-Cooper 变换，得到如下线性规划(LP)模型（产出导向）：

$1/E^* = \min v \cdot x_o$

s.t. $v \cdot X - w^{(1)} \cdot Z \ge 0$

$w^{(2)} \cdot Z - u \cdot Y \ge 0$

$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o}$

$u \cdot y_o = 1$ (9.30)

对偶包络模型

$1/E^* = \max \phi$

s.t. $X \cdot \lambda^{(1)} \le x_o$

$\mathbf{Z \cdot \lambda^{(1)} \ge z_o \cdot \tau}$

$\mathbf{Z \cdot \lambda^{(2)} \le z_o \cdot \tau}$

$Y \cdot \lambda^{(2)} \ge y_o \cdot \phi$

$\phi, \tau \in R$ (9.31)

基于中间流虚拟价值守恒的两阶段网络 DEA 模型具备以下特性：

1) 中间流柔性协同

放弃每个中间流价格必须相等的强外生约束，转而要求被评价 DMU 的所有中间流在系统内部流转时满足“总价值相等”，符合中间流内部结算不产生外部性的流转特征，使中间流权重（影子价格）可以在系统内部依据经典 DEA 的最优化原则确定。

$$\mathbf{w^{(1)} \cdot z_o = w^{(2)} \cdot z_o}$$

2) 纯粹的经典技术改进路径

完全采用经典的径向等比例技术改进路径，完整保留经典 DEA 模型的经济学意义。

$$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$$

$$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$$

3) 系统效率分解具有唯一性，并且分解出的阶段效率值是各自 PPS 下的技术效率。

9.1.3 两阶段网络 DEA 模型(CRS)的效率分解

两阶段网络 DEA 模型(9.27)，被评价 DMU 的整体网络效率 **E*** 等于阶段 1 模型(9.22)和阶段 2 模型(9.23)求解出的最优阶段效率 $E^{(1)}*$ 与 $E^{(2)}*$ 之积（**证明附后**）：

$$E^* = E^{(1)}* \times E^{(2)}*$$

9.1.4 中间流柔性协同机制

在网络 DEA 模型中，阶段独立确定的影子价格仅在阶段内部有效，其影子价格可以等效地整体放大或缩小。由于两个阶段的生产技术各自独立，当对系统整体效率进行评价时，需要一套协同机制来反映中间流在网络结构中的桥梁角色和中间流的流转规则。

柔性协同(Flexible Synergy)放弃了刚性协同(Rigid Synergy)的严苛约束（中间流价格一对一相等，$w^{(1)} = w^{(2)}$），放宽为“中间流在系统内部的生产端和消耗端总价值相等”($w^{(1)} \cdot z_o = w^{(2)} \cdot z_o$)。正是这项核心改变，实现了系统效率唯一性分解并使分解出的阶段效率与阶段独立技术效率保持一致。

从影子价格和生产函数理论视角剖析

在对偶规划与线性规划理论中，乘数可直接解释为各类投入、产出及中间流的内部影子价格(Shadow Prices)或边际估值。

刚性协同强制要求 $w^{(in)} = w^{(out)}$，为中间产品设置了一个跨阶段统一的“外生市场价格”。然而，阶段 1 与阶段 2 属于不同的生产技术，中间流在两个阶段的边际替代率(MRS)并不相同。强行要求中间流在两个阶段的影子价格一一相等，相当于用统一的线性切面去强行拟合两个独立的凸集前沿，必然导致拟合的前沿与两个阶段的前沿出现偏离，阶段效率被刚性协同约束所扭曲。

柔性协同仅约束中间流在系统内部转移时保持价值守恒，这赋予了 $w^{(1)}$和 $w^{(2)}$独立的自由度，使其能够分别精确匹配两阶段各自最优前沿面的法向量（即各自独立的局部最佳影子价格），从而使网络系统评估的阶段技术效率与该阶段独立评估的技术效率相一致。

从加权虚拟投入产出理论视角剖析

经典 DEA 的核心思想在于将多投入多产出聚合为标量型的虚拟投入(Virtual Input)与虚拟产出(Virtual Output)，并在虚拟产出/虚拟投入比值的约束下(<=1)，给被评价 DMU 完全的自由度来寻求最优权重。

中间流具有双重身份，既是阶段 1 的虚拟产出，同时又是阶段 2 的虚拟投入。刚性协同要求单个权重向量同时匹配两套投入产出结构，这限制了阶段各自的寻优自由度，导致模型无法实现阶

段的最优解。柔性协同解除了这一过度约束，仅施加中间产品虚拟聚合值相等约束。这既保证了阶段 1 向阶段 2 传递的虚拟价值在交界处既不凭空产生，也不无故消灭，完美契合中间产品仅存在于网络系统内部，不产生外部性的闭环特征，同时又保留了被评价 DMU 完全的自由度来寻求各阶段的最优权重，使各阶段在网络 DEA 架构下仍然能够实现各自的最优解。

表 9-1 刚性协同与柔性协同对中间流的约束比较

	刚性协同	柔性协同
乘数模型	刚性 中间流在两个阶段价格一对一相等 $w^{(1)} = w^{(2)}$	柔性 中间流在两个阶段总价值相等 $w^{(1)} \cdot z_o = w^{(2)} \cdot z_o$
包络模型	自由缩放 仅要求遵守流转逻辑 $z_i \cdot \lambda^{(2)} \le \hat{z}_i \le z_i \cdot \lambda^{(1)}, \forall_i$ 每项中间流独立自由缩放	径向缩放 遵守流转逻辑和中间流径向变动 $z_i \cdot \lambda^{(2)} \le z_{io} \cdot \tau \le z_i \cdot \lambda^{(1)}, \forall_i$ 或者表示为 $z_i \cdot \lambda^{(2)} \le \hat{z}_i \le z_i \cdot \lambda^{(1)}, \forall_i$ $\hat{z}_i = z_{io} \cdot \tau$ τ 为中间流缩放标量

* 当网络中只有一项中间流时，两种协同方式等价，$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o \rightarrow w^{(1)} = w^{(2)}$。

9.1.5 为什么两阶段网络模型的效率分解要使用乘积形式

两阶段系统效率分解的乘性结构避免了高效率阶段对低效率阶段的过度替代补偿。“两个阶段都不能落后”是两阶段网络系统的必然要求，效率乘积形式体现了典型的串联系统性质，任意子阶段的低效率（$E^{(t)*} < 1$）都会以乘法形式拉低系统整体效率。

举例说明：两家公司 A 和 B 用相同原料生产相同产品，A 公司在第一阶段用 100 件原料生产了 100 件中间产品，B 公司在第一阶段用 100 件原料生产了 10 件中间产品，与 A 相比，B 在第一阶段的效率是 0.1；A 公司在第二阶段利用 100 件中间产品生产了 100 件最终产品；B 公司在第二阶段利用 10 件中间产品生产了 9 件最终产品，与 A 相比，B 在第二阶段的效率是 0.9。乘积形式正确反映了这种串联连锁效应，B 的整体效率是 0.09，而不是算术平均形式得出的 0.5。

9.2 两阶段网络模型（规模报酬可变 VRS）

9.2.1 模型定义

1) 投入导向模型（截距归入产出端）：

阶段 1 模型：

$E^{(1)}* = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o)$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$ (9.32)

阶段 2 模型：

$E^{(2)}* = \max (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o)$

$s.t.\ w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$ (9.33)

两阶段网络模型：

$E* = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o) \cdot (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o) = \max (u \cdot y_o - u_o^{(2)}) / (v \cdot x_o)$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$

$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$

$(w^{(1)} \cdot z_o - u_o^{(1)}) = w^{(2)} \cdot z_o$ (9.34)

利用 Charnes-Cooper 变换，转换为线性规划。

$E* = \max (u \cdot y_o - u_o^{(2)})$

$s.t.\ v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$

$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$

$v \cdot x_o = 1$

$w^{(1)} \cdot z_o - u_o^{(1)} = w^{(2)} \cdot z_o$ (9.35)

对偶包络模型：

$E* = \min \theta$

$s.t.\ X \cdot \lambda^{(1)} \leq x_o \cdot \theta$

$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$

$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$

$Y \cdot \lambda^{(2)} \geq y_o$

$\mathbf{1} \cdot \lambda^{(1)} = \tau$

$\mathbf{1} \cdot \lambda^{(2)} = 1$

$\theta, \tau \in R$ (9.36)

2) 产出导向模型（截距归入投入端）：

阶段 1 模型：

$$E^{(1)*} = \max (w^{(1)} \cdot z_o - u_o^{(1)}) / (v \cdot x_o)$$
$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0 \qquad (9.37)$$

阶段 2 模型：

$$E^{(2)*} = \max (u \cdot y_o - u_o^{(2)}) / (w^{(2)} \cdot z_o)$$

$$\text{s.t. } w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0 \qquad (9.38)$$

两阶段网络模型：

$$E^* = \max (w^{(1)} \cdot z_o) / (v \cdot x_o + u_o^{(1)}) \cdot (u \cdot y_o) / (w^{(2)} \cdot z_o + u_o^{(2)}) = \max (u \cdot y_o) / (v \cdot x_o + u_o^{(1)})$$

$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$$
$$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$$
$$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o + u_o^{(2)} \qquad (9.39)$$

利用 Charnes-Cooper 变换，转换为线性规划。

$$1/E^* = \min (v \cdot x_o + u_o^{(1)})$$

$$\text{s.t. } v \cdot X - w^{(1)} \cdot Z + u_o^{(1)} \geq 0$$
$$w^{(2)} \cdot Z - u \cdot Y + u_o^{(2)} \geq 0$$
$$u \cdot y_o = 1$$
$$w^{(1)} \cdot z_o = w^{(2)} \cdot z_o + u_o^{(2)} \qquad (9.40)$$

对偶包络模型：

$$1/E^* = \max \phi$$

$$\text{s.t. } X \cdot \lambda^{(1)} \leq x_o$$

$$Z \cdot \lambda^{(1)} \geq z_o \cdot \tau$$

$$Z \cdot \lambda^{(2)} \leq z_o \cdot \tau$$

$$Y \cdot \lambda^{(2)} \geq \phi \cdot y_o$$

$$\mathbf{1} \cdot \lambda^{(1)} = 1$$

$$\mathbf{1} \cdot \lambda^{(2)} = \tau$$

$$\phi, \tau \in R \qquad (9.41)$$

9.2.2 效率分解

两阶段网络 DEA 模型(9.34)，被评价 DMU 的整体网络效率 $\mathbf{E^*}$ 等于阶段 1 模型(9.32)和阶段 2 模型(9.33)求解出的最优阶段效率 $\mathbf{E^{(1)*}}$ 与 $\mathbf{E^{(2)*}}$ 之积（与 CRS 模型原理相同，参见附录“CRS 两阶段网络 DEA 效率乘积分解证明”）：

$$\mathbf{E^* = E^{(1)*} \times E^{(2)*}}$$

9.3 广义网络 DEA 模型

9.3.1 模型定义

$E^* = \max \Sigma((u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) / (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)})) / N$

$s.t.\ v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + (\eta_3^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) = 0$ (9.42)

N 为网络的节点数量

η 用于表示 VRS 模型中截距项的归属：$\eta_1^{(t)} + \eta_2^{(t)} + \eta_3^{(t)} + \eta_4^{(t)} = 1, \eta \in \{0, 1\}, \forall t$

如果节点(t)没有投入 $x^{(t)}$：$\eta_1^{(t)} = 0$

如果节点(t)没有产出 $y^{(t)}$：$\eta_2^{(t)} = 0$

如果节点(t)没有中间流 $z^{(in)(t)}$：$\eta_3^{(t)} = 0$

如果节点(t)没有中间流 $z^{(out)(t)}$：$\eta_4^{(t)} = 0$

投入导向模型中截距项归入产出端：$\eta_2^{(t)} + \eta_4^{(t)} = 1$，如果节点(t)有产出 $y^{(t)}$，$\eta_2^{(t)} = 1$

产出导向模型中截距项归入投入端：$\eta_1^{(t)} + \eta_3^{(t)} = 1$，如果节点(t)有投入 $x^{(t)}$，$\eta_1^{(t)} = 1$

CRS 模型：η 均为 0

9.3.2 模型求解

先求解节点独立模型，获得每个节点的最优效率值 $E^{(t)*}$：

$E^{(t)*} = \max (u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) / (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)})$

$s.t.\ v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0$ (9.43)

广义网络模型的目标函数为分式相加形式，属于非线性规划，并且需要允许节点权重进行等比例放大，无法通过 Charnes-Cooper 变换实现线性化，需要采用 Dinkelbach 参数化方法求解。

参数化模型（投入导向，$\theta^{(t)}$是模型参数）：

$D^* = \max \Sigma((u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) - \theta^{(t)} \cdot (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)}))$

$s.t.\ v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} \geq 1, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + \eta_4^{(t)} \cdot u_o^{(t)}) = 0$ (9.44)

参数化模型（产出导向，$\phi^{(t)}$是模型参数）：

$D^* = \max \Sigma(\phi^{(t)} \cdot (u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) - (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)}))$

$s.t.\ v^{(t)} \cdot X^{(t)} + w^{(in)(t)} \cdot Z^{(in)(t)} - u^{(t)} \cdot Y^{(t)} - w^{(out)(t)} \cdot Z^{(out)(t)} + u_o^{(t)} \geq 0, \forall t$

$u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} \geq 1, \forall t$

$\Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + \eta_3^{(t)} \cdot u_o^{(t)}) = 0$ (9.45)

以投入导向为例，设置模型(9.44)的参数 $\theta^{(t)}$等于节点独立模型(9.43)得出的节点效率值 $E^{(t)*}$，求解模型，若模型的最优解目标函数值 $D^* = 0$，说明网络为**无损协同**；如果 $D^* < 0$ （D^*必定是 ≤ 0），说明网络为**有损协同**。

也可以求解参数化模型(9.44)的对偶模型：

$D^* = \max \Sigma(\alpha^{(t)})$

$s.t.\ X^{(t)}\cdot\lambda^{(t)} \le x_o^{(t)}\cdot(\theta^{(t)} - \alpha^{(t)}),\ \forall t$

$Z^{(in)(t)}\cdot\lambda^{(t)} \le z_o^{(in)(t)}\cdot(\theta^{(t)} - \alpha^{(t)} - \gamma),\ \forall t$

$Y^{(t)}\cdot\lambda^{(t)} \ge y_o^{(t)},\ \forall t$

$Z^{(out)(t)}\cdot\lambda^{(t)} \ge z_o^{(out)(t)}\cdot(1 - \gamma),\ \forall t$

$\gamma \in R,\ \alpha^{(t)} \ge 0$ (9.46)

VRS 模型：

如果节点(t)有产出 $y^{(t)}$： $\mathbf{1}\cdot\lambda^{(t)} = 1,\ \forall t$

如果节点(t)没有产出 $y^{(t)}$： $\mathbf{1}\cdot\lambda^{(t)} = 1 - \gamma,\ \forall t$

9.3.3 系统效率的计算方法

以投入导向模型为例：

1) 模型**无损协同**($D^* = 0$)

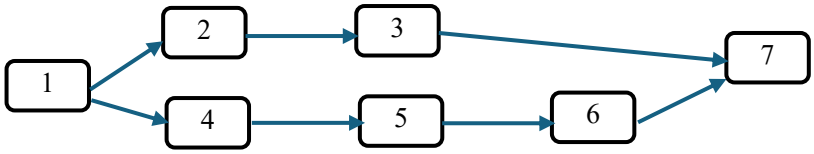
独立节点模型得出的节点效率值 $E^{(t)*}$与网络模型计算得出的节点效率值相等：

在模型可以实现**无损协同**时，网络模型的目标函数可以是节点效率的任意单调递增函数 $f(E^{(1)}, \dots, E^{(N)})$，节点独立最优解都可以组合构造出网络模型的最优解。据此，我们定义如下具有普适性的系统整体效率计算方法（幂乘积形式）：

$$E^* = \prod_{t=1}^{N}(E^{(t)*})^{\omega^{(t)}},\ 1 \le \sum\omega^{(t)} \le N$$

$\omega^{(t)}$反映每个节点在网络中的角色：**如果 $\omega^{(t)} = 1, \forall t$，则系统效率为所有节点效率之乘积；如果 $\omega^{(t)} = 1/N$，则系统效率为所有节点效率之几何平均；如果各节点的 $\omega^{(t)}$不相同，但 $\Sigma\omega^{(t)} = 1$，则系统效率为所有节点效率之加权几何平均。**

乘积形式适用于串联结构，任意子阶段的低效率（$E^{(t)*} < 1$）都会以乘法形式拉低系统整体效率；几何平均形式适用于并联结构；对于包含复杂串联和并联的网络结构，以下是一个确定节点权重的简单示例：



(1) 先按照串联关系为节点分配初始权重 1：

$\omega^{(1)} = 1$， $\omega^{(2)} = 1$， $\omega^{(3)} = 1$， $\omega^{(4)} = 1$， $\omega^{(5)} = 1$， $\omega^{(6)} = 1$， $\omega^{(7)} = 1$。

(2) 分支一（节点 2 和 3）与分支二（节点 4、5、6）是并联关系，对两个分支取几何平均（其内部节点的权重在原来的基础上除以 2）：

$\omega^{(1)} = 1$，$\omega^{(2)} = 0.5$，$\omega^{(3)} = 0.5$，$\omega^{(4)} = 0.5$，$\omega^{(5)} = 0.5$，$\omega^{(6)} = 0.5$，$\omega^{(7)} = 1$。

$\sum \omega^{(t)} = 4.5$，处于几何平均（$\sum \omega^{(t)} = 1$）和乘积（$\sum \omega^{(t)} = 7$）之间。

以上只是根据网络结构特征得出的权重，实证研究中，还可以结合节点对系统的重要程度确定最终的权重。

2) 模型**有损协同**($D^* < 0$)

如果 $D^* < 0$，则说明网络模型为有损协同，有损协同的节点效率值为

$E^{(t)-} = (u^{(t)*} \cdot y_o^{(t)} + w^{(out)(t)*} \cdot z_o^{(out)(t)} - u_o^{(t)*}) / (v^{(t)*} \cdot x_o^{(t)} + w^{(in)(t)*} \cdot z_o^{(in)(t)}) \leq E^{(t)*}$，

如果求解的是对偶模型(9.46)，则 $E^{(t)-} = E^{(t)*} - \alpha^{(t)*} \leq E^{(t)*}$。

有损协同的系统效率值为

$$E^- = \prod_{t=1}^{N} (E^{(t)-})^{\omega^{(t)}},\ 1 \leq \sum \omega^{(t)} \leq N$$

有损协同得出的节点效率值**不具有唯一性**。

3) 有损协同的解决方法

一种解决方法是综合运用“权重比值约束模型（又称为锥比率 DEA，约束锥 DEA，保证域 DEA）”来阻止一项或多项中间流的权重为 0，恰当地设置权重比值约束，可以实现无损协同。

另一种方法是采用后面介绍的“基于全维有效面(FDEF)的广义网络 DEA”，该模型的所有权重均严格大于 0，只需满足一个简单的条件（充分非必要条件），就能够实现无损协同：网络存在只生产中间流不消耗中间流（只有 $z^{(out)}$没有 $z^{(in)}$）的节点，同时也存在不生产中间流只消耗中间流（只有 $z^{(in)}$没有 $z^{(out)}$）的节点。前面的两阶段网络模型是满足这个条件的。

9.3.4 与前面的两阶段模型的兼容性

前面介绍的两阶段网络模型是广义网络模型在两阶段结构时的特例，两阶段模型满足**无损协同**的充分条件，如果采用乘积方式计算系统整体效率，虽然二者的最优解（权重 v, u, w）不相同，但得出的节点效率和系统效率相同。

9.4 基于全维有效面(FDEF)的广义网络 DEA

经典 DEA 模型常被提到的一个缺陷是效率测量没有包含通常会在模型中出现的投入和产出松弛，基于这个原因出现了各类直接基于松弛测量效率的非径向模型。模型中投入和产出松弛问题并非经典 DEA 本身的缺陷，而是数据样本自身没有能够建立覆盖所有样本投影区域的强有效前沿，造成径向投影点落在与坐标轴平行的弱有效前沿上，如下图所示，DMU A 的投影点为 A_1，位于与纵坐标轴平行的弱有效前沿上，投入 x_2的松弛为 A_1B。各类非径向模型的投影点则位于强有效前沿 BC 段。非径向模型虽然解决了经典模型效率测量没有包含松弛的问题，但带来了新的问题：一是削弱了经典模型的对偶乘数模型的经济学意义，二是投影点往往并不符合逻辑，例如加性模型和 SBM 模型是以最不利于被评价 DMU 的“最远距离”来确定投影点。以下图所示情况为例，在加性模型和 SBM 模型中，投影点是距离 A 更远的 C，而不是更近的 B。

经典包络模型中出现的松弛对应于乘数模型中出现权重为 0。还有一种处理经典模型中松弛问题的方案是全维有效面(FDEF)。在示例图中，前沿的 BC 段是一个 FDEF，基于 BC 仿射包（Affine Hull, 在二维坐标中是直线，在多维空间中是超平面）的径向投影点为 A_2，比 C 更合理。

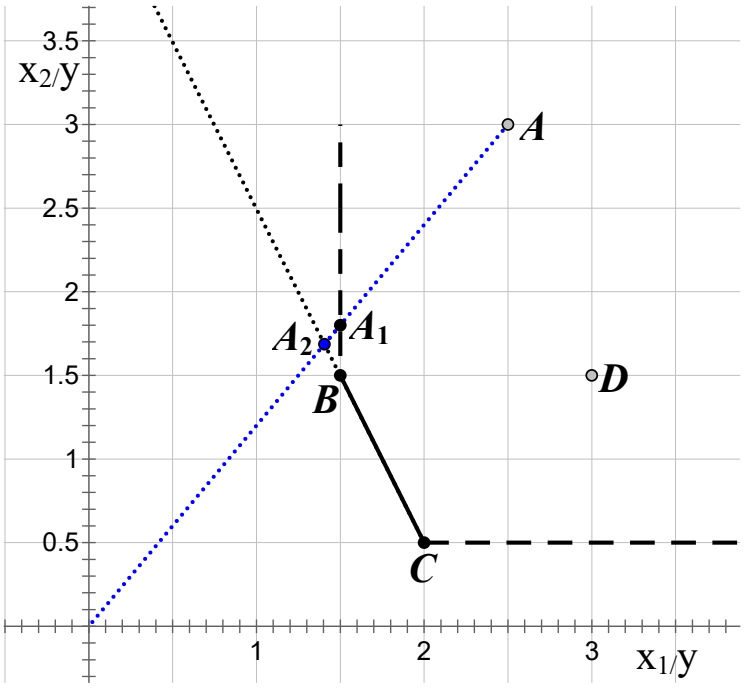


图 9-1 松弛处理方法示意图（两项投入一项产出 CRS 模型）

本节给出基于 FDEF 仿射包的广义网络 DEA 模型，可以消除经典模型中存在的松弛问题，包括投入和产出松弛，也包括中间流的松弛。中间流松弛的消失能够保证模型得出的第一阶段作为产出的中间流投影值完全等于第二阶段作为投入的中间流的投影值。

9.4.1 全维有效面(FDEF)简介

在数据包络分析(DEA)的凸包络几何理论中，全维有效面(Full-Dimensional Efficient Facets, FDEF)是刻画生产可能性集(PPS)前沿结构的核心概念之一。

1. 全维有效面(FDEF)的定义

在包含 m 个投入和 s 个产出的 DEA 模型中，生产可能性集 $PPS \subset \mathbb{R}^{m+s}$ 是一个多面体。如果该多面体边界上的某一有效超平面与其相交构成的有效面(Efficient Facet)，其几何维度恰好等于 $m+s-1$（即空间维度的最高可能自由度），则该面被称为**全维有效面(FDEF)**(Olesen & Petersen, 1996, 2003)。从乘数（对偶）角度来看，支撑 FDEF 的超平面法向量所对应的所有投入乘数 v 与产出乘数 u 均严格大于零(即 $v>0, u>0$)，不存在任何零权重。

2. 基于 FDEF 仿射包的 DEA 模型的优点

基于 FDEF 约束或识别构建的 DEA 模型具有以下显著优势：

彻底消除零权重问题：传统 DEA 模型常因乘数退化出现“零权重”现象（即某些投入或产出被赋予零权重），而 FDEF 保证了所有指标在边际评价中均发挥作用。

提供合理的经济学解释（边际替代率）：由于 FDEF 的法向量全为正值，由其导出的边际替代率(MRS)均为有限且非零的确定值，避免了低维有效面(LDEF)上边际替代率为 0 或无穷大的经济学失效问题。

优化投影与目标设置：将非有效决策单元(DMU)投影至 FDEF 超平面上，可以确保改进目标具备强有效性，彻底消除松弛问题。

3. 求解 DEA 模型样本中所有 FDEF 的方法

目前求解和识别多面体前沿上所有 FDEFs 的方法主要分为两类：

基于凸包(Convex Hull)的计算几何算法：利用计算几何中的顶点枚举与极面生成算法，对 DMU 顶点集构建凸包并提取维数为 $m+s-1$ 且法向量大于 0 的超平面(Olesen & Petersen, 2003)。

基于混合整数规划(MIP)的迭代搜索法：利用对偶理论，通过 MIP 搜寻所有权重严格大于零的超平面 (Olesen & Petersen, 2003; Jahanshahloo et al., 2007; Jahanshahloo et al., 2009; Olyaie, Roshdi, Jahanshahloo, & Asgharian, 2014)。

4. 样本不存在 FDEF 的现象与前沿严重退化

如果某一数据集在求解时不存在任何 FDEF，几何上意味着该样本的包络前沿完全由低维有效面(Lower-Dimensional Efficient Facets, LDEF)构成。这在理论与应用中具有明确的警示意义，表明前沿严重退化(Severe Frontier Degeneracy)，数据存在严重的病态数据(Ill-conditioned data)或维度灾难（即样本量相对于投入产出指标维度严重不足），导致生产前沿在高维空间中“折叠”或“塌陷”到低维子空间(Olesen & Petersen, 1996)。

一个有效的解决方法是引入足够数量的标杆 DMU 作为模型的参考集。

样本中存在一个 FDEF 并不意味着能够建立满意的评价模型，如果模型只有一个或极少量的 FDEF，被评价 DMU 有可能投影到严重远离 FDEF 的区域（位于 FDEF 的支撑超平面上，但严重远离 FDEF），得出不切实际的目标值。如何判断样本中的 FDEF 是否充足，是一个需要在理论和实践中继续探讨的问题。判断投影值（目标值）是否超出合理范围是判断 FDEF 是否充足的一个实用方法。

9.4.2 基于 FDEF 仿射包的广义网络 DEA 模型

基于 FDEF 仿射包的 DEA 模型就是把模型的权重限定为来自 FDEF 超平面集合，基于 FDEF 仿射包的广义网络 DEA 模型的表达式为：

$$E^* = \max \Sigma((u^{(t)} \cdot y_o^{(t)} + w^{(out)(t)} \cdot z_o^{(out)(t)} - (\eta_2^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) / (v^{(t)} \cdot x_o^{(t)} + w^{(in)(t)} \cdot z_o^{(in)(t)} + (\eta_1^{(t)} + \eta_3^{(t)}) \cdot u_o^{(t)})) / N$$

$$s.t.\ \Sigma(w^{(in)(t)} \cdot z_o^{(in)(t)} - w^{(out)(t)} \cdot z_o^{(out)(t)} + (\eta_3^{(t)} + \eta_4^{(t)}) \cdot u_o^{(t)}) = 0$$

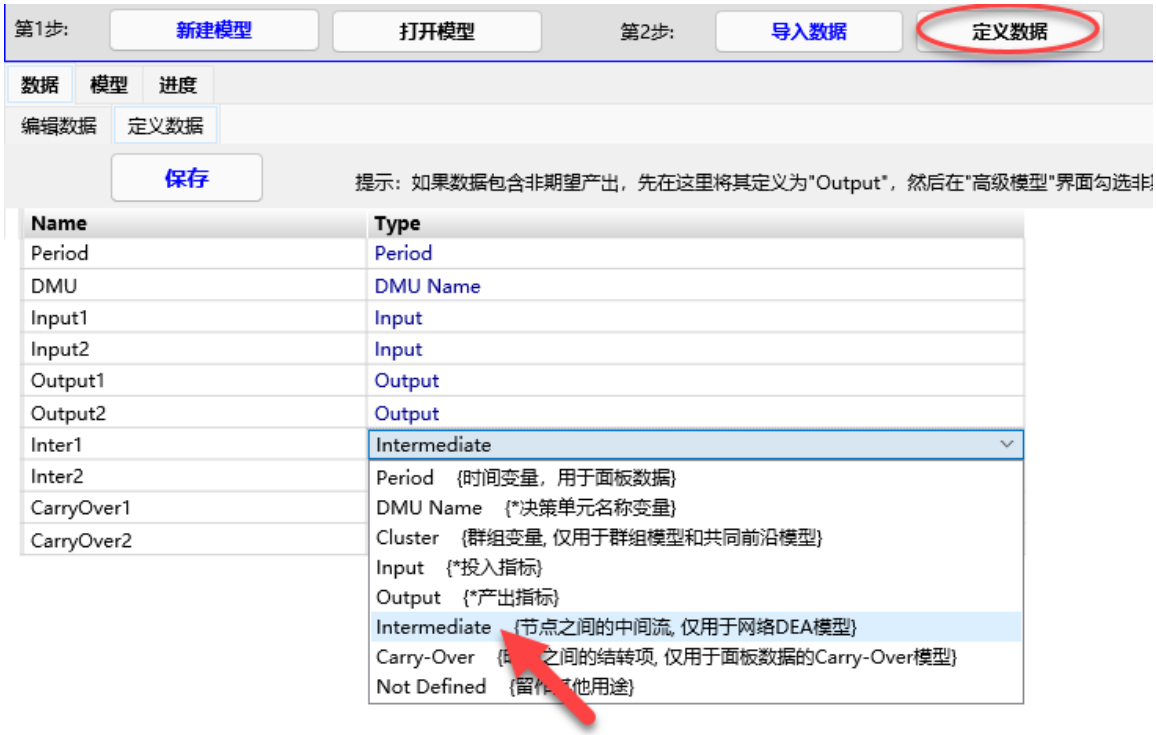
$$(v^{(t)}, u^{(t)}, w^{(in)(t)}, w^{(out)(t)}, u_o^{(t)}) \in F^{(t)} \qquad (9.47)$$

$F^{(t)}$表示阶段 t 的 FDEF 超平面集合。

基于 FDEF 仿射包的广义网络 DEA 模型，由于其权重全部严格大于 0，只需满足一个简单的条件（充分非必要条件），就能够实现无损协同：网络存在只生产中间流不消耗中间流（只有 $z^{(out)}$没有 $z^{(in)}$）的节点，同时也存在不生产中间流只消耗中间流（只有 $z^{(in)}$没有 $z^{(out)}$）的节点。

9.5 网络 DEA 的软件操作演示

9.5.1 定义数据



9.5.2 设置 DEA 模型的常规选项

生产函数类型（经典前沿生产函数或 FDEF 仿射包前沿生产函数）、模型导向和规模报酬。



9.5.3 设置其他模型选项

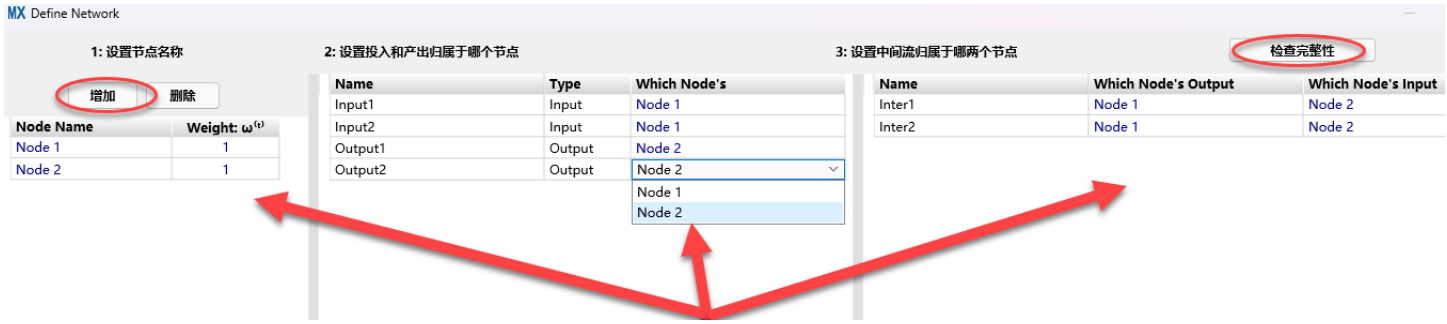
选择“**两阶段网络 DEA**”或“**广义网络 DEA**”。

网络 DEA 可以与所有“**面板数据模型**”进行组合，包括：Malmquist TFP 指数、HMB TFP 指数、窗口 DEA 和 Carry-Over DEA。



9.5.4 定义网络结构

如果选择“广义网络 DEA”，还需要定义网络结构。广义网络模型各节点的初始默认权重 $\omega^{(t)} = 1$，即系统效率为节点效率之乘积；如果需要系统效率为节点效率之几何平均，请设置各节点的权重之和等于 1，即 $\Sigma\omega^{(t)} = 1$。



9.5.5 输出结果

输出结果包括系统整体效率值、投入产出和中间流的权重、投入产出和中间流的投影值（目标值）、节点的效率值、节点的参考标杆及其系数(Lambda)等，VRS 模型还包括规模报酬常数项($u_0^{(t)}$)，广义网络模型还包括 D*（用于判断网络模型给出的节点效率值是否具有唯一性）。

附录：CRS 两阶段网络 DEA 效率乘积分解证明

两阶段网络 DEA 模型(9.27)，被评价 DMU 的整体网络效率 **E*** 等于阶段 1 模型(9.22)和阶段 2 模型(9.23)求解出的最优阶段效率 $\mathbf{E^{(1)*}}$ 与 $\mathbf{E^{(2)*}}$ 之积：

$$\mathbf{E^* = E^{(1)*} \times E^{(2)*}}$$

一、**证明** $\boldsymbol{e^* \le e^{(1)*} \times e^{(2)*}}$

设 $(v', w^{(1)r}, w^{(2)r}, u')$ 为系统整体模型的**任意一组可行解**。

利用网络模型的中间约束 $w^{(1)r} z_0 = w^{(2)r} z_0$，可将该可行解对应的目标函数值分解为：

$$\frac{u'y_0}{v'x_0} = \frac{w^{(1)r} z_0}{v'x_0} \cdot \frac{u'y_0}{w^{(2)r} z_0}$$

根据可行解的定义：

1. 向量对 $(v', w^{(1)r})$ 满足 $\frac{w^{(1)r} z_j}{v'x_j} \le 1\ (\forall j)$，因而 $(v', w^{(1)r})$ 是阶段 1 模型的一组可行解。根据 $e^{(1)*}$ 的最大值定义：

$$\frac{w^{(1)r} z_0}{v'x_0} \le e^{(1)*}$$

2. 向量对 $(w^{(2)r}, u')$ 满足 $\frac{u'y_j}{w^{(2)r} z_j} \le 1\ (\forall j)$，因而 $(w^{(2)r}, u')$ 是阶段 2 模型的一组可行解。根据 $e^{(2)*}$ 的最大值定义：

$$\frac{u'y_0}{w^{(2)r} z_0} \le e^{(2)*}$$

将上述两个不等式代入乘积分解式可得：

$$\frac{u'y_0}{v'x_0} \le e^{(1)*} \times e^{(2)*}$$

由于上式对网络模型可行域内的**任意**可行解均成立，因此对可行域上的最大值(即最优解)亦成立：

$$e^* \le e^{(1)*} \times e^{(2)*}$$

二、**证明** $\boldsymbol{e^* \ge e^{(1)*} \times e^{(2)*}}$

设 $(v^*, w^{(1)*})$ 为阶段 1 模型的一组**最优解**，满足：

$$e^{(1)*} = \frac{w^{(1)*} z_0}{v^* x_0}$$

设 $(w^{(2)*}, u^*)$ 为阶段 2 模型的一组**最优解**，满足：

$$e^{(2)*} = \frac{u^* y_0}{w^{(2)*} z_0}$$

构造一个正标量缩放因子：

$$\alpha = \frac{w^{(1)*} z_0}{w^{(2)*} z_0} > 0$$

利用阶段 2 模型分式规划的齐次性(即同比例放大分子分母权重不改变约束与比值)，对阶段 2 最优解进行线性缩放，构造新的权重向量：

$$\tilde{w}^{(2)} = \alpha w^{(2)*}, \quad \tilde{u} = \alpha u^*$$

验证组合解 $(v^*, w^{(1)*}, \tilde{w}^{(2)}, \tilde{u})$ 在网络模型中的可行性：

1. **阶段 1 约束**：$\frac{w^{(1)*} z_j}{v^* x_j} \le 1\ (\forall j)$ 显然满足。

2. **阶段 2 约束**：对于任意 j，有：

$$\frac{\tilde{u} y_j}{\tilde{w}^{(2)} z_j} = \frac{\alpha u^* y_j}{\alpha w^{(2)*} z_j} = \frac{u^* y_j}{w^{(2)*} z_j} \le 1$$

满足约束条件。

3. **中间流等式约束：**

$$\tilde{w}^{(2)}z_0 = (\alpha w^{(2)*})z_0 = \left(\frac{w^{(1)*}z_0}{w^{(2)*}z_0}\right)(w^{(2)*}z_0) = w^{(1)*}z_0$$

完美满足 $w^{(1)}z_0 = w^{(2)}z_0$ 的耦合约束。

因此，$(v^*, w^{(1)*}, \tilde{w}^{(2)}, \tilde{u})$ 是网络模型的一组**有效可行解**。

计算该可行解在网络模型中的目标函数值：

$$\frac{\tilde{u}y_0}{v^*x_0} = \frac{\alpha u^* y_0}{v^* x_0} = \frac{u^* y_0}{w^{(2)*}z_0} \cdot \frac{w^{(1)*}z_0}{v^* x_0} = e^{(2)*} \times e^{(1)*}$$

由于 e^* 是网络模型的全局最大值，而上述组合解仅为该模型可行域中的一个特定解，因此必有：

$$e^* \geq e^{(1)*} \times e^{(2)*}$$

三、结论

综合上界与下界的不等式：

$$e^* \leq e^{(1)*} \times e^{(2)*} \quad \text{且} \quad e^* \geq e^{(1)*} \times e^{(2)*}$$

根据实数大小关系的反对称性，必然推出：

$$e^* = e^{(1)*} \times e^{(2)*}$$

参考文献

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.