



# 基于数据包络分析的 保险应用研究



刘波 著



科学出版社

014033653

F222  
87

# 基于数据包络分析的 保险应用研究

刘波 著



中国保险学会教保基金资助课题 (编号: jiaobao2012-10)  
 辽宁省社会科学规划基金项目 (编号: L12BGL008)

科学出版社

北京



北航

C1721785

F222  
87

## 内 容 简 介

本书包括两大部分，第一部分由第1~5章构成，系统阐述数据包络分析的基本思想、方法、模型，效率和绩效的经济学基础以及数据包络分析与帕累托理论的关系。第二部分由第6~9章构成，将数据包络分析运用于社会医疗保险和商业保险的研究，包括新型农村合作医疗基金运行效率的研究、新型农村合作医疗补偿机制绩效的研究、产险业险种绩效的研究以及环境污染责任保险试点评价及优先发展区域与行业的研究。

本书可供保险业者、政府部门管理人员和相关领域的研究人员使用，也可供保险及相关专业的本科生、研究生和教师使用。

### 图书在版编目 (CIP) 数据

基于数据包络分析的保险应用研究 / 刘波著. —北京: 科学出版社, 2013

ISBN 978-7-03-037719-7

I. ①基… II. ①刘… III. ①统计数据—统计分析—应用—保险统计—研究 IV. ①F840.3

中国版本图书馆 CIP 数据核字 (2013) 第 120080 号

责任编辑: 马 跃 / 责任校对: 吴美艳  
责任印制: 阎 磊 / 封面设计: 蓝正设计

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京市文林印务有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2014 年 3 月第 一 版 开本: 720×1000 B5

2014 年 3 月第一次印刷 印张: 11 1/4

字数: 220 000

定价: 56.00 元

(如有印装质量问题, 我社负责调换)

## 序

刘波

刘波的《基于数据包络分析的保险应用研究》一书，是在其所承担相关课题的研究成果基础上整合而成的。该书对数据包络分析方法中与作者研究领域相关的内容做出了系统阐述和扩展，从效率和绩效两个视角，针对性地解释了数据包络分析基本模型和扩展模型的应用价值。在此基础之上，其对我国社会医疗保险领域的新型农村合作医疗、我国商业保险领域的产险业五大险种的效率与绩效问题，以及环境污染责任保险试点及重点区域与行业进行了应用研究。

除了较为系统地阐述数据包络分析基本理论与方法之外，该书还有三个研究焦点。其第一个焦点是我国的新型农村合作医疗(简称新农合)制度。新农合自其试点开始至今已有 10 年的时间，尽管已经基本实现了“全覆盖”，但由于政府补助力度与管理能力、医疗机构服务质量和农民接受性与缴费能力之间的矛盾冲突，新农合制度仍需很长一段时间才能得以完善。如何提高新农合基金的运行效率，建立符合各地区特点的补偿机制并使其发挥应有的作用是新农合管理部门面临的重要任务。该书借助数据包络分析方法，利用辽宁的新农合统计数据，分析了新农合基金的运行效率和补偿机制的绩效，并提出了改善效率和绩效的途径，尤其是将经济学中弹性的概念运用于补偿机制绩效的研究，从而找出了提高医疗费用补偿比的最优方案。

第二个焦点是我国产险业的险种结构。自“十一五”规划至“十二五”规划的几年间，我国产险业经历了“结构调整”与“结构优化升级”两个阶段，总结和回顾“十一五”期间产险业的“结构调整”成效，能够为“十二五”期间“结构优化升级”目标的实现提供方向。该书借助数据包络分析方法，利用一个城市各家产险公司的五大险种统计数据，分析了“结构调整”政策实施的成效，找出了影响各险种绩效的主要因素，并为改善险种绩效提出了建议。

第三个焦点是我国的环境污染责任保险制度。目前，环境污染责任保险尚处于试点阶段，因而数据的缺乏使众多学者的研究仅仅停留于理论探讨和国外经验借鉴的层面。该书选择了一个独特的研究角度，从中国 30 个省(自治区、直辖

市)、40个行业的现有环境污染物排放数据出发,利用数据包络分析的多风险综合评价方法,对环境污染责任保险试点分布做出了评价,对我国环境污染责任保险的重点发展省份和行业的选择提出了建议。

作者熟练地运用数据包络分析这一广受学者推崇的研究工具,对以上内容进行了深入研究,扩展了我国社会保险及商业保险应用研究的领域,因而是一部国内同类保险应用研究中不多见的、很有特色和新意的学术专著。

我高兴地看到作者在学术研究上的不断成长,因而欣然为本书作序。

艾洪德

2013年5月于东北财经大学

## 前 言

本书是作者所承担以下两项课题研究成果的组成部分：①2012年度中国保险学会教保基金资助课题，新型农村合作医疗补偿机制效率研究(编号：jiaobao2012-10)。②2012年度辽宁省社会科学规划基金项目，辽宁省新型农村合作医疗可持续发展研究(编号：L12BGL008)。

数据包络分析(data envelopment analysis, DEA)是数学、运筹学、数理经济学和管理学的一个新的交叉领域。它是由美国著名数学家和经济管理学家 A. Charnes 和 W. W. Cooper 等于 1978 年创建的，经过三十多年的发展，目前已经成为一种广受推崇的分析工具。众多学者借助这一分析工具，在数学、经济学、管理学、系统工程等领域从事研究，发表了大量的研究成果。这些成果表明，数据包络分析的适用范围极为广泛，摒弃了随机前沿法等生产前沿面测度的固有缺陷，因而在评估多投入、多产出的相对效率方面，数据包络分析是目前公认最有效的方法。不仅如此，数据包络分析还提供了种类繁多的效率测度且极易扩展，能使学者根据需要做出选择和改进，以卓有成效地服务于其研究目标。

作为一种强有力的分析工具，数据包络分析的基本理论已经形成一个较为完善并不断丰富的体系，有许多关于数据包络分析的经典优秀著作和论文问世，从而为本书作者在保险领域中的研究提供了坚实的基础。对此，作者的第一态度是学习，即从这些著作和论文中不断地汲取营养，并加以消化和吸收；作者的第二态度是学以致用，即努力将所学到的关于数据包络分析的理论与方法恰当运用于保险领域的研究。在研究过程中，随着对问题的探究的不断深入，又会发现自身对数据包络分析理论与方法的理解欠缺，就回过头来继续学习。这样一个螺旋式上升的过程使作者既加深了对数据包络分析理论与方法的理解，又提高了对数据包络分析模型的运用能力。另外，作者在一开始只能被动地使用其他学者编制的数据包络分析软件，因而在研究中受到很大制约。后来作者能够利用 Matlab 语言自行编制所需的程序，从而跨越了这个障碍，随着各种模型程序的不断丰富，作者对数据包络分析方法的使用就变得越来越得心应手。作者的第三个态度是学

无止境,随着数据包络分析理论体系的不断丰富完善和本人对其理解的不断加深,作者将继续虚心学习,潜心研究。如果读者能够从本书中发现疏漏之处并不吝赐教,作者将不胜感谢并虚心接受。

近几年来,作者关注的研究领域之一是新型农村合作医疗制度。本书的第6章和第7章分别展示了作者关于新型农村合作医疗基金运行效率和补偿机制绩效的研究成果。作者关注的研究领域之二是我国的财产保险业。本书的第8章展示了作者关于我国产险业五大险种(包括财产险、车辆险、船舶险、货运险和人身险)绩效的研究成果。作者关注的研究领域之三是我国的环境污染责任保险。本书的第9章展示了作者关于我国环境污染责任保险试点评价,以及优先发展区域和行业选择的研究成果。这些研究成果的获得都基于数据包络分析理论与方法的运用。正是得益于对数据包络分析方法的掌握,才使得作者的研究视角具有某种程度的独特性。如果这些研究成果能够对保险领域的研究起到一点扩展作用,能够对我国保险领域的研究做出一点贡献,则是作者感到非常欣慰和愉悦的事情。

本书包括两大部分,第一部分由第1~5章构成,系统阐述数据包络分析的基本思想、方法、模型,效率和绩效的经济学基础以及数据包络分析与帕累托理论的关系。

第1章综述数据包络分析方法基本原理。首先介绍数据包络分析的基本要素,包括决策单元、输入与输出、生产可能集、生产函数与规模报酬等。其次通过“单输入—单输出”“双输入—单输出”和“单输入—双输出”三种简单的情形来解释数据包络分析的基本思想。最后介绍数据包络分析效率指标及指标分解。

第2章介绍数据包络分析的基本模型。首先分别介绍两种最经典的模型: $C^2R$ 模型和 $BC^2$ 模型。其次介绍数据包络分析效率指标及另一种指标分解。最后给出 Matlab 程序和例题。

第3章介绍数据包络分析的扩展模型。首先介绍只有输出指标的数据包络分析模型,包括输出极大模型、输出极小模型和输出极大极小模型。其次介绍自由处置变量模型和非自由处置变量模型。

第4章介绍数据包络分析的面板数据模型。首先介绍 Malmquist 指数,其次介绍基于面板数据的扩展模型。

第5章阐述效率与绩效的经济学原理及应用。首先分别阐述效率和绩效的经济学基础,并指出效率和绩效与帕累托最优之间的关系。其次阐述数据包络分析关于效率和绩效的度量与帕累托最优的等价性。

第二部分由第6~9章构成,将数据包络分析运用于社会医疗保险和商业保险的研究。每一章都按科学研究成果的基本写作方式(研究背景、文献回顾、研究方法、研究指标与数据来源、研究过程、研究结论与建议)相对独立成章,使读者形成整体的认识。

第6章是关于新型农村合作医疗基金运行效率的研究。首先利用 $C^2R$ 和 $BC^2$ 模型计算新型农村合作医疗基金的效率(包括纯技术效率、规模效率和技术效率)并进行比较分析。其次进行Malmquist指数度量与分析。

第7章是关于新型农村合作医疗补偿机制绩效的研究。借助数据包络分析模型,首先度量辽宁各县(区、市)新型农村合作医疗补偿机制的绩效并进行比较分析,其次分析影响补偿机制绩效的因素,最后通过引入经济学中关于弹性的基本思想,对提高医疗费用补偿比的最优方案做出选择。

第8章是关于产险业险种绩效的研究。借助数据包络分析模型,以大连的统计数据为例,分别度量各家产险公司五大险种的绩效,以评价我国保险业“十一五”规划期间险种“结构调整”政策的成效,并分析了影响险种绩效的因素。

第9章是关于环境污染责任保险试点评价及优先发展区域与行业的研究。借助数据包络分析模型,借助来自《中国统计年鉴》的数据,对全国各经济区域(包括东部、中部和西部)的环境污染风险进行了评价,通过风险等级划分,分析了我国环境污染责任保险试点分布的合理性。之后,对我国40个行业的环境污染风险进行了评价并进行了等级划分。在此基础上,为我国优先发展环境污染责任保险的区域和行业提出了建议。

在本书的研究过程中,先后有多名教师、业界人士和研究生加入了作者的研究团队。作者攻读博士学位期间的导师孙刚教授提供了许多指导。作者的同事刘晓梅教授、董普教授、张抗私教授贡献了许多宝贵意见和建议。辽宁省财政厅的李鑫和大连市财政局的艾广青在调研和研究数据上提供了支持。大连保监局的一些工作人员提供了许多宝贵意见和建议。作者的研究生任旭、曹云波、王靓、马明煜、张云霞、权娜、罗雯、岳琳、刘露、夏文硕、陈晓云、解婷馨、孙书衡参与了课题的研究工作,宁静在研究数据方面提供了支持,他们有的已经成为大学教师,有的已经在金融、保险行业工作,有的即将毕业,都对课题的研究投入了满腔热情和自己的聪明才智。在本书的撰写过程中,作者的研究生杜宪玲、宗会苓、陈琦、岳鹏伟、张露伶、翟晓程、王贵、袁耀杰、王希月参与了文献检索、制图、文字编辑、校对等工作。本书的撰写与出版还得到了东北财经大学金融学院院长邢天才教授的大力支持。在此,作者一并表示感谢。

作者

2014年1月于东北财经大学



# 目 录

第 1 章 数据包络分析方法基本原理	1
1.1 引言	1
1.2 数据包络分析的基本思想	9
1.3 输入导向效率和输出导向效率	12
参考文献	14
第 2 章 数据包络分析的基本模型	15
2.1 $C^2R$ 模型	15
2.2 $BC^2$ 模型	21
2.3 技术效率、纯技术效率和规模效率	25
2.4 Matlab 程序与例题	26
参考文献	28
附录	29
第 3 章 数据包络分析的扩展模型	32
3.1 只有输出的数据包络分析模型	32
3.2 自由处置变量模型和非自由处置变量模型	35
参考文献	39
第 4 章 数据包络分析的面板数据模型	40
4.1 Malmquist 指数	40
4.2 基于面板数据的扩展模型	42
参考文献	46
第 5 章 效率与绩效的经济学原理及应用	47
5.1 效率	47
5.2 绩效	49
5.3 帕累托最优	50
5.4 数据包络分析与帕累托最优	51

参考文献 .....	54
<b>第 6 章 新型农村合作医疗基金运行效率研究 .....</b>	<b>57</b>
6.1 新型农村合作医疗发展历程 .....	57
6.2 文献回顾 .....	65
6.3 模型 .....	69
6.4 指标体系与数据来源 .....	70
6.5 效率分析 .....	71
6.6 Malmquist 指数分析 .....	75
6.7 结论与建议 .....	79
参考文献 .....	80
<b>第 7 章 新型农村合作医疗补偿机制绩效研究 .....</b>	<b>82</b>
7.1 新型农村合作医疗补偿机制状况 .....	82
7.2 文献回顾 .....	86
7.3 模型 .....	89
7.4 指标体系与数据来源 .....	91
7.5 绩效分析 .....	92
7.6 影响因素分析 .....	102
7.7 补偿比最优方案的选择 .....	106
7.8 结论与建议 .....	111
参考文献 .....	114
附录 .....	116
<b>第 8 章 产险业险种绩效研究 .....</b>	<b>119</b>
8.1 产险业发展历程 .....	119
8.2 文献回顾 .....	121
8.3 研究方法 .....	127
8.4 指标体系与数据来源 .....	129
8.5 险种绩效评价 .....	130
8.6 结论与建议 .....	144
参考文献 .....	145
<b>第 9 章 环境污染责任保险试点及优先发展区域与行业研究 .....</b>	<b>149</b>
9.1 环境污染责任保险发展历程 .....	149
9.2 文献回顾 .....	150
9.3 研究方法 .....	155
9.4 指标体系与数据来源 .....	159
9.5 区域风险评价 .....	159

9.6 行业风险评价 .....	163
9.7 结论与建议 .....	165
参考文献 .....	167

# 第1章

## 数据包络分析方法基本原理

### 1.1 引言

数据包络分析(data envelopment analysis, DEA)的理论可以追溯到1957年,经济学家 Michael Farrell 基于生产效率测度思想进行了具有开创性的研究工作。他在分析英国农业生产率时引入了前沿生产函数的概念,其核心思想是根据一组已知样本的投入和产出指标,定义由样本中所有有效率的单位(即一定生产要素投入下能够达到最大产出的样本单位)构成的生产前沿面(production frontiers)或效率边界,使其他所有样本值落入该边界之内,根据每个样本与边界的距离状况,计算该样本的效率值,这一效率值是相对于生产前沿面的效率值。生产前沿面是指特定生产技术条件下各种生产要素投入的配置可能生产的最大产出所描述的生产可能性边界。生产前沿面的生产函数称为前沿生产函数或边界生产函数(frontier production function)。

生产前沿面的研究涉及是否构造前沿生产函数。参数法沿袭了传统的生产函数估计思想,它首先根据生产性质确定具体的生产函数,然后利用回归分析或其他方法估计位于生产前沿面上的函数参数。关于参数法,目前主要有三种,包括随机前沿法(stochastic frontier approach, SFA)、自由分布法(distribution-free approach, DFA)和厚前沿法(thick frontier approach, TFA)。

参数法的优点是稳定性强,能够检验参数估计值的有效性。其缺点是:①事先确定的生产函数形式可能与实际情形不符;②尽管便于处理单一投入、单一产出的情形,但对多投入、多产出的情形难以处理;③仅能得出效率值的大小,而无法指出改进的方向。

非参数法摒弃了传统的生产函数估计思想,在分析生产单元的效率时,在一定的生产效率性标准下,利用所观察的大量生产点数据,确定位于生产前沿面的相对有效点,从而构造效率前沿面。非参数法主要包括数据包络分析和自由处理包(free disposal hull, FDH)。数据包络分析方法将凸性生产可能集的最佳运营观测值连接成为一个分段线性组合,由此构造其前沿面(称为 DEA 前沿面),因此无须明确生产函数形式。自由处理包方法是数据包络分析方法的特例,它不将连接 DEA 前沿面各顶点连线上的点集视为效率前沿,其生产可能集仅由 DEA 前沿面顶点和这些顶点内部的自由排列组成,因而自由处理包方法得到的效率值常常高于数据包络处理方法得到的效率均值。

与参数法相比,非参数法的优点是:①无须知道前沿生产函数的形式,从而避免与实际不相符的可能性;②便于处理多投入、多产出的情形;③不仅能够得出效率值,还能指出一个缺乏效率的企业与“最佳实务”(best practice)相比的改进方向。然而,非参数法也具有一些缺点:①没有考虑运气成分、数据问题或其他计量问题所致随机误差;②不便于检验结果的显著性;③对效率值的估计偏低、离散程度较大;④一般忽略价格因素,仅仅关注技术层面的无效率性(即过多投入或过少产出)。

尽管以数据包络分析方法为代表的非参数法存在一些缺点,但在评估多投入、多产出相对效率方面,数据包络分析方法是目前公认的最有效的方法。由于本书旨在介绍数据包络分析方法,并将其应用于保险的效率研究,因而本章会重点介绍这种方法的基本内容。

数据包络分析是数学、运筹学、数理经济学和管理科学的一个新的交叉领域。它是由美国著名数学家和经济管理学家 A. Charnes 和 W. W. Cooper 等于 1978 年始建。数据包络分析使用数学规划模型评价具有多个输入、特别是多个输出的“部门”或“策略业务单位”(strategic business unit)之间的相对有效性。借助数据包络分析方法和模型,可以确定生产前沿面的结构、特征和构造方法,因此又可将其看做一种非参数的统计估计方法。数据包络分析是在经济中用来研究具有多输入、多输出的边界生产函数的有力工具,因而可用来研究与边界生产函数有关的问题,如狭义技术进步率、生产率、预测值、规模报酬分析、资金分配问题、最小成本问题和最大收益问题等。数据包络分析已经成为经济、管理、决策分析等领域中的一个重要方法。

数据包络分析可以看做一种统计分析的新方法,它是根据一组关于“输入—输出”的观察值来估计有效生产前沿面的。在有效性的评价方面,除了数据包络分析方法以外,还有其他的一些方法,但是那些方法几乎仅限于单输出的情况。相比之下,数据包络分析方法处理多输入特别是多输出问题的能力是具有绝对优势的。使用数据包络分析对决策单元(decision making unit, DMU)进行效率评

价时,可以得到很多在经济学中具有深刻经济含义和背景的管理信息,而且数据包络分析方法与输入、输出指标的量纲选取无关。因此,它比其他一些方法(包括统计方法)优越,在各行各业得到广泛的应用,并得以迅速发展。由于在评价有效性方面具有优势,因此数据包络分析方法吸引了众多的应用者。

数据包络分析的第一个模型—— $C^2R$ 模型产生于1978年。从生产函数角度看,这一模型是用来研究具有多个输入、特别是具有多个输出的“生产部门”同时为“规模有效”与“技术有效”的十分理想且卓有成效的方法。Banker等(1984)给出了一个称为 $BC^2$ 的模型,Charnes等(1985)给出了一个称为 $C^2GS^2$ 的模型,这两个模型是用来研究生产部门之间的“技术有效性”的。就数据包络分析研究领域而言,这些模型以及新的模型正在不断得到完善和进一步发展,而且在被不断运用到实际工作中。

上述模型都可以看做处理具有多个输入越小越好和多个输出越大越好的多目标决策问题的方法。已经证明的一个结论是:数据包络分析的有效性与其相应的多目标规划问题的帕累托(Pareto)有效解是等价的。

数据包络分析方法直接使用多个决策单元的投入、产出数据,运用数学规划模型,对决策单元做出综合评价,得出决策单元效率性的综合数量指标,识别有效率的决策单元,定量地指出其他决策单元缺乏效率的原因和程度,从而获得反映系统状态的各种管理信息。

数据包络分析方法对具有多项投入指标和多项产出指标项的复杂系统有很强的适用性,它以决策单元各项输入、输出的权重为优化变量,从最有利于决策单元的角度进行评价,从而避免了确定各方都认为合理公正的权重这一棘手问题,且增加了评价的客观性与科学性。数据包络分析可用于评价决策单元在各时期不同技术经济条件下的发展是否相对有效率,分析缺乏效率的原因,指出调整以达到有效率状态的途径与程度。更为重要的是,通过数据包络分析评价,可获得评价对象发展过程中的有效率点。由此,一切有效率点构成了评价对象发展的有效轨迹,从而可获得经验有效生产前沿面。这将为确定评价对象的生产函数、预测有效面的变化、分析技术进步等奠定基础。数据包络分析模型的基础是数学规划,其理论基础牢固、方法简捷。因此,与其他方法相比,数据包络分析方法具有较大的优良性,对于社会、科技、经济等不同领域中的评价问题也具有很强的应用价值。

数据包络分析的上述优点可归纳为:

(1)无需构造一个确定的基本生产函数和估计函数参数系数,可避免人为的因采取错误的函数形式得出错误的结论,并且不需要投入品的价格。

(2)对于复杂系统,其子系统种类较多,并且各子系统指标之间难以比较。数据包络分析可以通过数学规划方式客观产生权重,并且有效处理投入或产出单

位不一致的问题。

(3)数据包络分析方法无须事先给定输入、输出权向量,而是先将其看做变向量,然后在分析过程中根据某种原则来确定,排除了很多主观的因素,因而具有很强的客观性。

(4)决策单元的各输入、输出之间关系极其复杂,而数据包络分析方法不必确定这种函数关系就可以得出每个决策单元综合效率的数量指标,据此确定有效决策单元,并对非有效的决策单元分析原因,为主管部门提供管理信息,进一步调整决策单元投入规模的正确方向和程度。

(5)数据包络分析可以有效处理定性与定量投入产出指标的问题,即可处理比率尺度与顺序尺度数据兼容性,且较少受观察值多寡之限制。从全局、整体的角度利用数据,从而避免了通常分散处理指标的片面性。

(6)数据包络分析综合模型不受加权、排序等外界人为因素的影响,具有较强的操作性。

### 1.1.1 决策单元

一个经济系统或一个生产过程可以看成是一个单元在一定的可能范围内,通过投入一定数量生产要素并产生一定数量产品的活动,虽然这种活动的具体内容各不相同,但其目的都是尽可能地使这一活动取得最大效益。由于产出是决策的结果,所以这样的单元被称为决策单元。因此,可以认为每个决策单元都代表或表现出一定的经济意义,其基本特点是具有一定的输入和输出,并且在将输入转化成输出的过程中,努力实现自身的决策目标。在数据包络分析中,使用较多的是同类型的决策单元。所谓同类型的决策单元,是指具有以下三个特征的集合:①具有相同的目标和任务;②具有相同的外部环境;③具有相同的输入和输出指标。

### 1.1.2 输入与输出

按照系统的语言,“投入”通常被称为“输入”,而“产出”通常被称为“输出”。这样,一个决策单元就是将确定数目的“输入”转化为确定数目的“输出”的实体。

根据研究目的的不同,即使同一个决策单元,其“输入”与“输出”也有所不同。例如,在研究高校学生培养质量时,学生的人数和平均成绩可以被视为输入;但在研究高校发展水平时,学生人数和平均成绩可以被视为输出。在利用数据包络分析进行效率研究时,必须根据研究的需要来确定输入和输出指标。这表明输入和输出指标不可以随意确定,而是由研究视角决定。

### 1.1.3 生产可能集

对于决策单元, 在一项生产或经营活动中, 如果其输入变量为  $x = (x_1, x_2, \dots, x_m)^T$ , 输出变量为  $y = (y_1, y_2, \dots, y_s)^T$ , 我们就可以简单地以  $(x, y)$  来表示这个决策单元的整个生产活动。

**定义 1.1** 称集合

$$T = \{(x, y) \mid \text{投入 } x \text{ 能够获得 } y\}$$

为所有可能的生产活动构成的生产可能集。

**定义 1.2** 称由  $(x_j, y_j)$ ,  $j=1, 2, \dots, n$  组成的集合

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

为参考集。

为了便于进行研究, 通常假设生产可能集满足以下四个公理:

**公理 1:** 凸性。对于任意的  $(x, y) \in T$  和  $(x', y') \in T$ , 以及  $\mu \in [0, 1]$ , 有

$$\mu(x, y) + (1-\mu)(x', y') \in T$$

凸性表明  $T$  是一个凸集。

**公理 2:** 锥性。如果  $(x, y) \in T$ , 且  $k \geq 0$ , 则:

$$k(x, y) = (kx, ky) \in T$$

其含义是, 若将原输入的规模同时扩大  $k$  倍, 总是可能得到原输出的  $k$  倍。

**公理 3:** 无效性。对于  $(x, y) \in T$

$$\text{如果 } x' \geq x, \text{ 则 } (x', y) \in T$$

$$\text{如果 } y' \geq y, \text{ 则 } (x, y') \in T$$

其含义是, 以在原生产活动的基础之上增加投入或减少产出的方式进行生产总是可能的。

**公理 4:** 最小性。生产可能集  $T$  是满足公理 1 至公理 3 的所有集合的交集。

对于满足公理 1 至公理 4 的生产活动, 如果其观察值为  $(x_j, y_j)$ ,  $j=1, 2, \dots, n$ , 可以得到

$$T = \left\{ (x, y) \mid k \sum_{j=1}^n \mu_j x_j \leq x, k \sum_{j=1}^n \mu_j y_j \geq y, \sum_{j=1}^n \mu_j = 1, \mu_j \geq 0, j = 1, 2, \dots, n, k > 0 \right\}$$

在上式中, 令  $k\mu_j = \lambda_j$ ,  $j=1, 2, \dots, n$ , 则上式变形为

$$T_1 = \left\{ (x, y) \mid \sum_{j=1}^n \lambda_j x_j \leq x, \sum_{j=1}^n \lambda_j y_j \geq y, \lambda_j \geq 0, j = 1, 2, \dots, n \right\} \quad (1.1)$$

当生产活动不满足公理 2 时, 式(1.1)变为

$$T_2 = \left\{ (x, y) \mid \sum_{j=1}^n \lambda_j x_j \leq x, \sum_{j=1}^n \lambda_j y_j \geq y, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, j = 1, 2, \dots, n \right\} \quad (1.2)$$



在研究生产活动的规模报酬时，还会考虑的生产可能是

$$T_3 = \left\{ (x, y) \mid \sum_{j=1}^n \lambda_j x_j \leq x, \sum_{j=1}^n \lambda_j y_j \geq y, \sum_{j=1}^n \lambda_j \leq 1, \lambda_j \geq 0, j = 1, 2, \dots, n \right\} \quad (1.3)$$

**例题 1.1** 表 1.1 为三个决策单元的单输入、单输出形式的生产活动。

表 1.1 单输入  $x$ 、单输出  $y$  情形

DMU	A	B	C
输入 $x$	1	2	3
输出 $y$	2	3	1

这三个决策单元的生产可能集  $T_1$  如图 1.1 所示，它通过射线  $OA$  将所有生产可能点包络进来。因而，经过原点和  $A$  的射线称为包络面。由于这个包络面是由观察数据点  $A$ 、 $B$  和  $C$  得到的，因此这个包络面称为数据包络面。

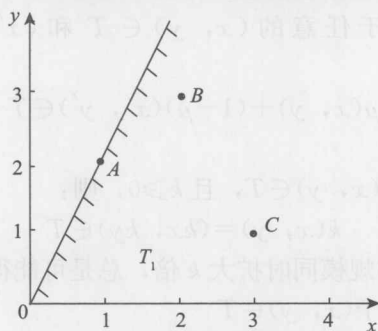


图 1.1 生产可能集  $T_1$  图例

这三个决策单元的生产可能集  $T_2$  如图 1.2 所示，它通过经过点  $A$  的垂直线、线段  $AB$  和经过点  $B$  的水平线，将所有生产可能点包络进来，由此形成数据包络面。

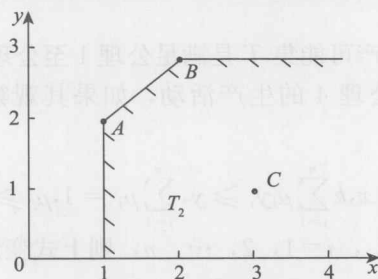


图 1.2 生产可能集  $T_2$  图例

这三个决策单元的生产可能集  $T_3$  如图 1.3 所示，它通过线段  $OA$ 、 $AB$  和经过点  $B$  的水平线，将所有生产可能点包络进来，由此形成数据包络面。