
Data Envelopment Analysis

Second Edition

DATA ENVELOPMENT ANALYSIS

**A Comprehensive Text with Models,
Applications, References
and DEA-Solver Software**

Second Edition

WILLIAM W. COOPER

University of Texas at Austin, U.S.A.

LAWRENCE M. SEIFORD

University of Michigan, U.S.A.

KAORU TONE

National Graduate Institute for Policy Studies, Japan

William W. Cooper
University of Texas, USA

Lawrence M. Seiford
University of Michigan, USA

Kaoru Tone
National Graduate Institute for Policy Studies, Japan

Library of Congress Control Number: 2006932712

ISBN-10: 0-387-45281-8 (HB) ISBN-10: 0-387-45283-4 (e-book)
ISBN-13: 978-0387-45281-4 (HB) ISBN-13: 978-0387-45283-8 (e-book)

Printed on acid-free paper.

© 2007 by Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if the are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com

To

Ruth, Bev and Michiko

Contents

List of Tables	xvii
List of Figures	xxi
Preface	xxv
1. GENERAL DISCUSSION	1
1.1 Introduction	1
1.2 Single Input and Single Output	2
1.3 Two Inputs and One Output Case	6
1.4 One Input and Two Outputs Case	8
1.5 Fixed and Variable Weights	12
1.6 Summary and Conclusion	13
1.7 Problem Supplement for Chapter 1	15
2. BASIC CCR MODEL	21
2.1 Introduction	21
2.2 Data	22
2.3 The CCR Model	23
2.4 From a Fractional to a Linear Program	23
2.5 Meaning of Optimal Weights	25
2.6 Explanatory Examples	25
2.6.1 Example 2.1 (1 Input and 1 Output Case)	26
2.6.2 Example 2.2 (2 Inputs and 1 Output Case)	27
2.7 Illustration of Example 2.2	30
2.8 Summary of Chapter 2	32
2.9 Selected Bibliography	33
2.10 Problem Supplement for Chapter 2	34
3. CCR MODEL AND PRODUCTION CORRESPONDENCE	41
3.1 Introduction	41
3.2 Production Possibility Set	42
3.3 The CCR Model and Dual Problem	43
3.4 The Reference Set and Improvement in Efficiency	47

3.5	Theorems on CCR-Efficiency	48
3.6	Computational Aspects of the CCR Model	50
3.6.1	Computational Procedure for the CCR Model	50
3.6.2	Data Envelopment Analysis and the Data	52
3.6.3	Determination of Weights (=Multipliers)	52
3.6.4	Reasons for Solving the CCR Model Using the Dual (DLP_o)	52
3.7	Example	53
3.8	The Output-Oriented Model	58
3.9	An Extension of the Two Phase Process in the CCR Model	60
3.10	Discretionary and Non-Discretionary Inputs	63
3.11	Summary of Chapter 3	68
3.12	Notes and Selected Bibliography	68
3.13	Related DEA-Solver Models for Chapter 3	70
3.14	Problem Supplement for Chapter 3	71
4.	ALTERNATIVE DEA MODELS	87
4.1	Introduction	87
4.2	The BCC Models	89
4.2.1	The BCC Model	91
4.2.2	The Output-oriented BCC Model	93
4.3	The Additive Model	94
4.3.1	The Basic Additive Model	94
4.3.2	Translation Invariance of the Additive Model	97
4.4	A Slacks-Based Measure of Efficiency (SBM)	99
4.4.1	Definition of SBM	100
4.4.2	Interpretation of SBM as a Product of Input and Output Inefficiencies	101
4.4.3	Solving SBM	101
4.4.4	SBM and the CCR Measure	103
4.4.5	The Dual Program of the SBM Model	104
4.4.6	Oriented SBM Models	105
4.4.7	A Weighted SBM Model	105
4.4.8	Decomposition of Inefficiency	106
4.4.9	Numerical Example (SBM)	106
4.5	A Hybrid Measure of Efficiency (Hybrid)	106
4.5.1	A Hybrid Measure	107
4.5.2	Decomposition of Inefficiency	109
4.5.3	Comparisons with the CCR and SBM Models	110
4.5.4	An Illustrative Example	111
4.6	Russell Measure Models	112
4.7	Summary of the Basic DEA Models	114
4.8	Summary of Chapter 4	116
4.9	Notes and Selected Bibliography	117
4.10	Appendix: Free Disposal Hull (FDH) Models	117
4.11	Related DEA-Solver Models for Chapter 4	119
4.12	Problem Supplement for Chapter 4	120

5.	RETURNS TO SCALE	131
5.1	Introduction	131
5.2	Geometric Portrayals in DEA	134
5.3	BCC Returns to Scale	136
5.4	CCR Returns to Scale	138
5.5	Most Productive Scale Size	143
5.6	Further Considerations	147
5.7	Relaxation of the Convexity Condition	150
5.8	Decomposition of Technical Efficiency	152
5.8.1	Scale Efficiency	152
5.8.2	Mix Efficiency	154
5.8.3	An Example of Decomposition of Technical Efficiency	155
5.9	An Example of Returns to Scale Using a Bank Merger Simulation	156
5.9.1	Background	156
5.9.2	Efficiencies and Returns to Scale	156
5.9.3	The Effects of a Merger	159
5.10	Summary	162
5.11	Additive Models	162
5.12	Multiplicative Models and "Exact" Elasticity	165
5.13	Summary of Chapter 5	170
5.14	Appendix: FGL Treatment and Extensions	171
5.15	Related DEA-Solver Models for Chapter 5	172
5.16	Problem Supplement for Chapter 5	173
6.	MODELS WITH RESTRICTED MULTIPLIERS	177
6.1	Introduction	177
6.2	Assurance Region Method	178
6.2.1	Formula for the Assurance Region Method	178
6.2.2	General Hospital Example	181
6.2.3	Change of Efficient Frontier by Assurance Region Method	183
6.2.4	On Determining the Lower and Upper Bounds	184
6.3	Another Assurance Region Model	185
6.4	Cone-Ratio Method	186
6.4.1	Polyhedral Convex Cone as an Admissible Region of Weights	186
6.4.2	Formula for Cone-Ratio Method	187
6.4.3	A Cone-Ratio Example	188
6.4.4	How to Choose Admissible Directions	189
6.5	An Application of the Cone-Ratio Model	189
6.6	Negative Slack Values and Their Uses	194
6.7	A Site Evaluation Study for Relocating Japanese Government Agencies out of Tokyo	196
6.7.1	Background	196
6.7.2	The Main Criteria and their Hierarchy Structure	197
6.7.3	Scores of the 10 Sites with respect to the 18 Criteria	198
6.7.4	Weights of the 18 Criteria by the 18 Council Members (Evaluators)	199
6.7.5	Decision Analyses using Averages and Medians	201

6.7.6	Decision Analyses using the Assurance Region Model	201
6.7.7	Evaluation of "Positive" of Each Site	202
6.7.8	Evaluation of "Negative" of Each Site	202
6.7.9	Uses of "Positive" and "Negative" Scores	203
6.7.10	Decision by the Council	203
6.7.11	Concluding Remarks	204
6.8	Summary of Chapter 6	205
6.9	Notes and Selected Bibliography	206
6.10	Related DEA-Solver Models for Chapter 6	206
6.11	Problem Supplement for Chapter 6	207
7.	NON-DISCRETIONARY AND CATEGORICAL VARIABLES	215
7.1	Introduction	215
7.2	Examples	217
7.3	Non-controllable, Non-discretionary and Bounded Variable Models	219
7.3.1	Non-controllable Variable (NCN) Model	219
7.3.2	An Example of a Non-Controllable Variable	220
7.3.3	Non-discretionary Variable (NDSC) Model	222
7.3.4	Bounded Variable (BND) Model	224
7.3.5	An Example of the Bounded Variable Model	224
7.4	DEA with Categorical DMUs	227
7.4.1	An Example of a Hierarchical Category	227
7.4.2	Solution to the Categorical Model	228
7.4.3	Extension of the Categorical Model	229
7.5	Comparisons of Efficiency between Different Systems	231
7.5.1	Formulation	231
7.5.2	Computation of Efficiency	232
7.5.3	Illustration of a One Input and Two Output Scenario	232
7.6	Rank-Sum Statistics and DEA	233
7.6.1	Rank-Sum-Test (Wilcoxon-Mann-Whitney)	234
7.6.2	Use of the Test for Comparing the DEA Scores of Two Groups	235
7.6.3	Use of the Test for Comparing the Efficient Frontiers of Two Groups	236
7.6.4	Bilateral Comparisons Using DEA	236
7.6.5	An Example of Bilateral Comparisons in DEA	237
7.6.6	Evaluating Efficiencies of Different Organization Forms	238
7.7	Summary of Chapter 7	240
7.8	Notes and Selected Bibliography	240
7.9	Related DEA-Solver Models for Chapter 7	240
7.10	Problem Supplement for Chapter 7	242
8.	ALLOCATION MODELS	257
8.1	Introduction	257
8.2	Overall Efficiency with Common Prices and Costs	258
8.2.1	Cost Efficiency	258
8.2.2	Revenue Efficiency	260
8.2.3	Profit Efficiency	260
8.2.4	An Example	261

8.3	New Cost Efficiency under Different Unit Prices	262
8.3.1	A New Scheme for Evaluating Cost Efficiency	262
8.3.2	Differences Between the Two Models	264
8.3.3	An Empirical Example	265
8.3.4	Extensions	267
8.4	Decomposition of Cost Efficiency	269
8.4.1	Loss due to Technical Inefficiency	269
8.4.2	Loss due to Input Price Inefficiency	270
8.4.3	Loss due to Allocative Inefficiency	271
8.4.4	Decomposition of the Actual Cost	271
8.4.5	An Example of Decomposition of Actual Cost	272
8.5	Summary of Chapter 8	272
8.6	Notes and Selected Bibliography	273
8.7	Related DEA-Solver Models for Chapter 8	274
8.8	Problem Supplement for Chapter 8	276
9.	DATA VARIATIONS	283
9.1	Introduction	283
9.2	Sensitivity Analysis	283
9.2.1	Degrees of Freedom	283
9.2.2	Algorithmic Approaches	284
9.2.3	Metric Approaches	284
9.2.4	Multiplier Model Approaches	287
9.3	Statistical Approaches	291
9.4	Chance-Constrained Programming and Satisficing in DEA	298
9.4.1	Introduction	298
9.4.2	Satisficing in DEA	298
9.4.3	Deterministic Equivalents	299
9.4.4	Stochastic Efficiency	302
9.5	Summary of Chapter 9	304
10.	SUPER-EFFICIENCY MODELS	309
10.1	Introduction	309
10.2	Radial Super-efficiency Models	310
10.3	Non-radial Super-efficiency Models	313
10.3.1	Definition of Non-radial Super-efficiency Measure	314
10.3.2	Solving Super-efficiency	315
10.3.3	Input/Output-Oriented Super-efficiency	316
10.3.4	An Example of Non-radial Super-efficiency	316
10.4	Extensions to Variable Returns-to-Scale	317
10.4.1	Radial Super-efficiency Case	317
10.4.2	Non-radial Super-efficiency Case	318
10.5	Summary of Chapter 10	319
10.6	Notes and Selected Bibliography	319
10.7	Related DEA-Solver Models for Chapter 10	319
10.8	Problem Supplement for Chapter 10	320

11. EFFICIENCY CHANGE OVER TIME	323
11.1 Introduction	323
11.2 Window Analysis	324
11.2.1 An Example	324
11.2.2 Application	324
11.2.3 Analysis	326
11.3 Malmquist Index	328
11.3.1 Dealing with Panel Data	328
11.3.2 Catch-up Effect	329
11.3.3 Frontier-shift Effect	329
11.3.4 Malmquist Index	330
11.3.5 The Radial MI	331
11.3.6 The Non-radial and Slacks-based MI	333
11.3.7 The Non-radial and Non-oriented MI	336
11.3.8 Scale Efficiency Change	337
11.3.9 Illustrative Examples for Model Comparisons	338
11.3.10 Concluding Remarks	344
11.4 Summary of Chapter 11	345
11.5 Notes and Selected Bibliography	345
11.6 Related DEA-Solver Models for Chapter 11	345
12. SCALE ELASTICITY AND CONGESTION	349
12.1 Introduction	349
12.2 Scale Elasticity in Production	350
12.3 Congestion	353
12.3.1 Strong Congestion	354
12.3.2 Weak Congestion	357
12.3.3 Summary of Degree of Scale Economies and Congestion	360
12.4 Illustrative Examples	360
12.4.1 Degree of Scale Economies and Strong Congestion	360
12.4.2 Weak vs. Strong Congestion	361
12.5 Summary of Chapter 12	362
12.6 Notes and Selected Bibliography	363
12.7 Related DEA-Solver Models for Chapter 12	364
12.8 Problem Supplement for Chapter 12	364
13. UNDESIRABLE OUTPUTS MODELS	367
13.1 Introduction	367
13.2 An SBM with Undesirable Outputs	368
13.2.1 An Undesirable Output Model	368
13.2.2 Dual Interpretations	369
13.2.3 Returns-to-scale (RTS) Issues	370
13.2.4 Imposing Weights to Inputs and/or Outputs	370
13.3 Non-separable 'Good' and 'Bad' Output Model	371
13.4 Illustrative Examples	374
13.4.1 Separable Bad Outputs Models	374
13.4.2 An Example with Both Separable and Non-separable Inputs/Outputs	375

13.5 Comparisons with Other Methods	376
13.6 Summary of Chapter 13	378
13.7 Related DEA-Solver Models for Chapter 13	378
14. ECONOMIES OF SCOPE AND CAPACITY UTILIZATION	381
14.1 Introduction	381
14.2 Economies of Scope	382
14.2.1 Definition	382
14.2.2 Checking for Economies of Scope	382
14.2.3 Checking a Virtual Merger	386
14.2.4 Comparisons of Business Models	387
14.2.5 Illustrative Examples	388
14.3 Capacity Utilization	390
14.3.1 Fixed vs. Variable Input Resources	390
14.3.2 Technical Capacity Utilization	391
14.3.3 Price-Based Capacity Utilization Measure	392
14.3.4 Long-Run and Short-Run Capacity Utilization	395
14.3.5 Illustrative Examples	396
14.4 Summary of Chapter 14	401
14.5 Notes and Selected Bibliography	401
14.6 Related DEA-Solver Models for Chapter 14	402
14.7 Problem Supplement for Chapter 14	402
15. A DEA GAME	405
15.1 Introduction	405
15.2 Formulation	406
15.3 Coalition and Characteristic Function	409
15.4 Solution	411
15.4.1 Coalition and Individual Contribution	411
15.4.2 The Shapley Value	411
15.5 DEA min Game	414
15.6 Summary of Chapter 15	415
15.7 Notes and Selected Bibliography	416
15.8 Problem Supplement for Chapter 15	416
16. MULTI-STAGE USE OF PARAMETRIC AND NON-PARAMETRIC MODELS	423
16.1 Introduction	423
16.2 OLS Regressions	423
16.3 Modification and Extensions	424
16.4 Stochastic Frontier Analysis and Composed Error Models	426
16.5 DEA and Regression Combinations	427
16.6 Multi-stage DEA-Regression Combinations and its Application to Japanese Banking	428
16.6.1 Introduction	428
16.6.2 The Multistage Framework	430
16.6.3 An Application to Japanese Banking	433

16.6.4 Discussion	438
16.6.5 Summary of This Case Study	439
16.7 Summary of Chapter 16	439
16.8 Notes and Selected Bibliography	439
Appendices	443
A—Linear Programming and Duality	443
A.1 Linear Programming and Optimal Solutions	443
A.2 Basis and Basic Solutions	443
A.3 Optimal Basic Solutions	444
A.4 Dual Problem	445
A.5 Symmetric Dual Problems	446
A.6 Complementarity Theorem	447
A.7 Farkas' Lemma and Theorem of the Alternative	448
A.8 Strong Theorem of Complementarity	449
A.9 Linear Programming and Duality in General Form	451
B—Introduction to DEA-Solver	454
B.1 Platform	454
B.2 Installation of DEA-Solver	454
B.3 Notation of DEA Models	454
B.4 Included DEA Models	456
B.5 Preparation of the Data File	456
B.5.1 The CCR, BCC, IRS, DRS, GRS, SBM, Super-Efficiency, Scale Elasticity, Congestion and FDH Models	456
B.5.2 The AR Model	457
B.5.3 The ARG Model	458
B.5.4 The NCN and NDSC Models	459
B.5.5 The BND Model	460
B.5.6 The CAT, SYS and Bilateral Models	460
B.5.7 The Cost and New-Cost Models	461
B.5.8 The Revenue and New-Revenue Models	462
B.5.9 The Profit, New-Profit and Ratio Models	462
B.5.10 The Window and Malmquist Models	462
B.5.11 The Hybrid Model	463
B.5.12 Weighted SBM Model	464
B.5.13 The Bad Outputs Model	465
B.5.14 The Non-separable Outputs Model	465
B.6 Starting DEA-Solver	466
B.7 Results	466
B.8 Data Limitations	473
B.8.1 Problem Size	473
B.8.2 Inappropriate Data for Each Model	474
B.9 Sample Problems and Results	475
B.10 Summary	475
B.10.1 Models that Require Numbers to be Supplied through Keyboard	475

B.10.2 Summary of Headings to Inputs/Outputs	475
C–Bibliography	477

Index	479
-------	-----

Index	483
-------	-----

List of Tables

1.1	Single Input and Single Output Case	3
1.2	Efficiency	5
1.3	Two Inputs and One Output Case	6
1.4	One Input and Two Outputs Case	8
1.5	Hospital Case	12
1.6	Comparisons of Fixed vs. Variable Weights	13
1.7	Optimal Weights for Hospitals <i>A</i> and <i>B</i>	18
2.1	Example 2.1	26
2.2	Results of Example 2.1	27
2.3	Example 2.2	28
2.4	Results of Example 2.2	29
3.1	Primal and Dual Correspondences	44
3.2	Example 3.1	53
3.3	Results of Example 3.1	58
3.4	Problem for Phase III Process	61
3.5	CCR-Score, Reference Set, Slacks and % Change	62
3.6	State-mandated Excellence Standards on Student Outcomes	67
3.7	Non-Discretionary Inputs	67
3.8	Worksheets Containing Main Results	71
3.9	Data and Scores of 5 Stores	78
3.10	Optimal Weights and Slacks	79
3.11	CCR-projection in Input and Output Orientations	81
4.1	Primal and Dual Correspondences in BCC Model	92
4.2	Data and Results of Example 4.1	96
4.3	Data and Results of CCR and SBM	107
4.4	A Comparison: Hybrid, CCR and SBM	111
4.5	Measures of Inefficiency: (<i>Hybrid</i>)	112
4.6	Summary of Model Characteristics	115
4.7	Decomposition of Efficiency Score	129
5.1	Decomposition of Technical Efficiency	155
5.2	Data of 11 Regional and 9 City Banks*	157

5.3	Efficiencies and Returns to Scale	158
5.4	Weights, Slacks and Projections	160
5.5	Efficiency of Projected and Merged Banks	160
5.6	Results of Input-oriented/Output-oriented BCC Cases	174
6.1	Data for 14 Hospitals	181
6.2	Efficiency and Weight of 14 Hospitals by CCR Model	182
6.3	Efficiency and Weight of 14 Hospitals with Assurance Region Method	182
6.4	Efficiency of 14 Hospitals by CR (Cone-Ratio) and CCR Models	188
6.5	Number of Bank Failures (through 10-31-88)	191
6.6	Inputs and Outputs	192
6.7	CCR and Cone-Ratio Efficiency Scores (1984, 1985)*	193
6.8	Printout for Cone-Ratio CCR Model - Interstate Bank of Fort Worth, 1985.	195
6.9	Scores (S_{ij}) of 10 Sites (A-J) with respect to 18 Criteria (C1-C18)	198
6.10	Statistics of Weights assigned the 18 Criteria (C1-C18) by 18 Council Members	200
6.11	Averages and Medians of Scores of the 10 Sites	201
7.1	Data for Public Libraries in Tokyo	220
7.2	Efficiency of Libraries by CCR and NCN	221
7.3	Data of 12 Japanese Baseball Teams in 1993	225
7.4	Projection of Attendance by CCR and Bounded Models	226
7.5	Categorization of Libraries	228
7.6	Nine DMUs with Three Category Levels	230
7.7	Comparisons of Stores in Two Systems	234
7.8	Comparisons of Two Systems	234
7.9	Example of Bilateral Comparisons	238
8.1	Sample Data for Allocative Efficiency	261
8.2	Efficiencies	261
8.3	Comparison of Traditional and New Scheme	265
8.4	Data for 12 Hospitals	266
8.5	New Data Set and Efficiencies	266
8.6	Decomposition of Actual Cost	272
9.1	Data for a Sensitivity Analysis	289
9.2	Initial Solutions	290
9.3	Results of 5% Data Variations	290
9.4	OLS Regression Estimates without Dummy Variables	294
9.5	Stochastic Frontier Regression Estimates without Dummy Variables	295
9.6	OLS Regression Estimates without Dummy Variables on DEA-efficient DMUs	296
9.7	Stochastic Frontier Regression Estimates without Dummy Variables on DEA-efficient DMUs	297

10.1	Test Data	311
10.2	Andersen-Petersen Ranking*	312
10.3	Non-radial Super-efficiency	317
10.4	Data for Super-efficiency	320
10.5	Super-efficiency Scores under Variable RTS	321
10.6	Super-efficiency Scores under Constant RTS	321
11.1	Window Analysis: 56 DMUs in U.S. Army Recruitment Battalions 3 Outputs - 10 Inputs	325
11.2	Example 1	339
11.3	Input-oriented Scores	339
11.4	Catch-up, Frontier-shift and Malmquist Index	339
11.5	Example 2	340
11.6	Example 3	342
11.7	Comparisons	343
11.8	Example 4	343
11.9	Results by the Non-oriented Non-radial Model	344
12.1	Example 1	361
12.2	Example 2	362
12.3	BCC-O Results	362
12.4	Congestion Results	363
13.1	Separable Bad Outputs Case: Data set	374
13.2	Separable Bad Outputs Case: Results	375
13.3	Non-Separable Inputs/Outputs Case: Data Set	376
13.4	Non-Separable Inputs/Outputs Case: Decomposition of Inefficiency	377
14.1	Data for Groups 1 and 2 as Specialized and Group 3 as Diversified Firms	389
14.2	Twenty Virtual Diversified Firms	390
14.3	Efficiency Score and Degree of Economies of Scope	390
14.4	Data for 12 Hospitals	397
14.5	Technical Capacity Utilization	398
14.6	Price-Based Data Set	398
14.7	Profits and Losses	399
14.8	Comparisons of Current and Maximum Profits	399
15.1	Score Matrix	405
15.2	Division of Reward based on Fixed Weights	406
15.3	Optimal Rewards with Optimal Weights	407
15.4	Normalized Score Matrix	409
15.5	Coalition and Characteristic Function (1)	410
15.6	Coalition and Characteristic Function (2)	411
15.7	Each Member's Marginal Contribution to Coalitions	412
15.8	The Shapley Value	413
15.9	Division of Reward based on Shapley Value	413
15.10	Single Players' Values for the Min Game Case	416
15.11	Coalitions' Values for the Min Game Case (1)	416

15.12	Coalitions' Values for the Min Game Case (2)	417
15.13	The Shapley Value for the Min Game Case	418
15.14	Data of 3 Shops	418
15.15	Single Players' Values of the Market Arcade Game	419
15.16	Coalitions' Values of the Market Arcade Game	419
15.17	The Shapley Value of the Market Arcade Game	419
15.18	Characteristic Function Values subject to the AR Constraints	420
15.19	The Shapley Value for the AR Case	420
16.1	Stochastic Frontier Estimation Results	436
16.2	Comparison of the Initial and Final Efficiency Scores	437
A.1	Symmetric Primal-Dual Problem	447
A.2	General Form of Duality Relation	452
B.1	Window Analysis by Three Adjacent Years	469
B.2	Sample Data of Scale Elasticity	471
B.3	Models Requiring Numbers Through Keyboard	475
B.4	Headings to Inputs/Outputs	476

List of Figures

1.1	Comparisons of Branch Stores	3
1.2	Regression Line vs. Frontier Line	4
1.3	Improvement of Store <i>A</i>	5
1.4	Two Inputs and One Output Case	7
1.5	Improvement of Store <i>A</i>	8
1.6	One Input and Two Outputs Case	9
1.7	Improvement	9
1.8	Excel File “HospitalAB.xls”	17
2.1	Example 2.1	27
2.2	Example 2.2	30
2.3	Region <i>P</i>	31
2.4	The Case of DMU <i>A</i>	32
2.5	A Scatter Plot and Regression of DEA and Engineering Efficiency Ratings	36
2.6	Envelopment Map for 29 Jet Engines	37
2.7	A Comparative Portrayal and Regression Analysis for 28 Engines (All Data Except for Engine 19)	38
3.1	Production Possibility Set	43
3.2	Example 3.1	57
4.1	Production Frontier of the CCR Model	88
4.2	Production Frontiers of the BCC Model	88
4.3	The BCC Model	90
4.4	The Additive Model	95
4.5	Translation in the BCC Model	98
4.6	Translation in the Additive Model	98
4.7	FDH Representation	117
4.8	Translation in the CCR Model	121
5.1	Returns to Scale	132
5.2	Returns to Scale: One Output-One Input	134
5.3	Most Productive Scale Size	140
5.4	The IRS Model	150

5.5	The DRS Model	151
5.6	The GRS Model	152
5.7	Scale Efficiency	153
5.8	Merger is Not Necessarily Efficient.	161
6.1	Efficiency Scores with and without Assurance Region Constraints	183
6.2	Assurance Region	184
6.3	Convex Cone generated by Two Vectors	187
6.4	The Hierarchical Structure for the Capital Relocation Problem	199
6.5	Positives and Negatives of the 10 Sites	204
7.1	One Input Exogenously Fixed, or Non-Discretionary	217
7.2	DMUs with Controllable Category Levels	230
7.3	Comparisons between Stores using Two Systems	233
7.4	Bilateral Comparisons—Case 1	237
7.5	Bilateral Comparisons—Case 2	237
7.6	Case 1	249
7.7	Case 2	249
7.8	Controllable Input and Output	251
7.9	Categorization (1)	252
7.10	Categorization (2)	252
8.1	Technical, Allocative and Overall Efficiency	258
8.2	Decomposition of Actual Cost	273
9.1	Stable and Unstable DMUs	285
9.2	A Radius of Stability	286
10.1	The Unit Isoquant Spanned by the Test Data in Table 10.1	311
11.1	Catch-up	329
11.2	Scale Efficiency Change	338
11.3	Three DMUs in Two Periods	341
12.1	Scale Elasticity	350
12.2	Congestion	353
14.1	Diversified and Specialized Firms	383
14.2	Diversified Firms D and Virtual Diversified Firms V	384
14.3	Firms with Two Different Scopes	387
14.4	Price-Based Capacity Utilization	395
14.5	Decomposition of the Maximum Profit	400
14.6	Price-Based Capacity Utilization with a Variable Output	403
16.1	Downward Trend in Mean Measured Efficiency by Static DEA Model	430
16.2	Comparison of Mean Efficiency between Initial and Final Estimates	438
B.1	Sample.xls in Excel Sheet	457
B.2	Sample-AR.xls in Excel Sheet	458
B.3	Sample-ARG.xls in Excel Sheet	459
B.4	Sample-NCN (NDSC).xls in Excel Sheet	459
B.5	Sample-BND.xls in Excel Sheet	460

B.6	Sample-CAT.xls in Excel Sheet	461
B.7	Sample-Cost(New-Cost).xls in Excel Sheet	461
B.8	Sample-Revenue(New-Revenue).xls in Excel Sheet	462
B.9	Sample-Window(Malmquist).xls in Excel Sheet	463
B.10	Sample-Hybrid.xls in Excel Sheet	464
B.11	Sample-Weighted SBM.xls in Excel Sheet	464
B.12	Sample-BadOutput.xls in Excel Sheet	465
B.13	Sample-NonSeparable.xls in Excel Sheet	466
B.14	A Sample Score Sheet	467
B.15	A Sample Graph2	468
B.16	Variations through Window	469
B.17	Variations by Term	470
B.18	Sample Malmquist Index	470
B.19	Malmquist Index (Graph)	471
B.20	Worksheets “Elasticity”	472
B.21	Worksheets “Congestion”	472
B.22	Worksheets “Decomposition”	473

Preface

Preface to the Second Edition

1. Introduction

This is the second edition of *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software* the book which we published with Kluwer Academic Publishers, 2000. The present version corrects and reproduces the 9 chapters in that book, where the corrections were made partly in response to reader comments and suggestions—for which we are grateful. It also incorporates Chapter 10 on “Super Efficiency Models” from our paperback volume, *Introduction to Data Envelopment Analysis and Its Uses* (Springer-Science + Business Media, Inc., 2006). The present revision adds 6 new chapters which reflect relatively recent developments that greatly extend the power and scope of DEA and lead to new directions for research and possible uses.

2. Added Chapters

In a sense, these added new chapters represent advanced treatments of DEA that we make available for further use after the first 10 chapters have been suitably covered. We therefore explain their scope and potential in summary fashion as follows:

Chapter 11 covers efficiency changes over time. It extends from “window analysis” (covered in the first edition) and extends to the “Malmquist Index.” Window analysis may be regarded as a generalization of the “moving average analyses” that are taught in introductory statistics courses. The presentation here is effected in a manner that makes it possible to simultaneously examine trends in the efficiency evaluations and the stability of efficiency scores, as different Decision Making Units (=DMUs)¹ are substituted for use in the evaluations.

The “Malmquist Index,” also covered in this chapter, is used to evaluate DMU performances between two points in time. It differs from other index numbers in its ability to distinguish between improved possibilities of perfor-

mance (e.g., due to technological change) and the degree to which each of the evaluated entities has taken advantage of these possibilities, in both periods.

Much of the past work on Malmquist Indexes has been deficient because of failure of the measures that were used to reflect all of the inefficiencies. Using recent work by Tone, and others, this deficiency is here overcome and a new more comprehensive version of these Malmquist measures is provided in the present text.

Because this portion of the book represents an advanced text, we do not generally provide problems to accompany and elaborate the developments. We do, however, supply explanatory materials that facilitates use of the DEA solver code to help in uses of these developments.

Chapter 12 turns to scale elasticity and congestion. These topics were covered separately, and in a more introductory manner, in the earlier chapters. Here the two, scale elasticity and congestion, are combined and extended for joint use in potential new applications. This is important especially in public sector applications. An example occurs in state owned enterprises in China where an excessive number of employees is evident with, in fact, some of these firms paying employees to stay away and not appear for work, presumably to avoid the negative effects of congestions on outputs. In such cases it is necessary to consider both scale changes and congestion possibilities and, simultaneously, provide for employee allocation and reallocation between such enterprises.

In many DEA applications it is necessary to consider undesirable as well as desirable outputs. This is the topic of Chapter 13. In some cases the desirable and undesirable outputs are separate but in other cases they are not separate. In Chapter 13 this is illustrated with an example involving electric utilities in which the use of fossil fuels to generate electric power (good output) is accompanied by smoke and other emissions (bad outputs). This is in contrast to hydroelectric power which generates electricity without such emissions. There are also cases in between, as when hydroelectric generation is accompanied by steam generators in either a “stand by” or “auxiliary” capacity.

Chapter 14 deals with “economies of scope.” This brings in the topic of whether it is more efficient to produce two or more products in the same plant (or company) or whether efficiency is improved by producing them in separate plants or companies. Thus, the issue is whether diversification is more efficient than specialization in the products to be produced.

As originally formulated, economies-of-scope analyses, required access to internal company records. Developments in the DEA literature, however, modified this so that access to data obtainable from public records suffices. This opens the prospect of bringing these approaches to bear on issues of public policy (as well as internal management).

An example application of economies of scope to electric utility firms is used in Chapter 14 where some firms specialize in generation and others specialize in distribution of electric power. This is complemented by a third group of companies that engage in both generation and distribution so that the wanted comparisons can be made.

Another part of Chapter 14 deals with “capacity utilization” in a manner that makes it possible to determine whether it is beneficial to expand the capacity for producing some of the outputs without worsening other outputs. Potential uses are then illustrated with applications to hospital data in which the number of doctors represents fixed capacity and the number of nurses is variable.

Chapter 15 turns to “n-person cooperative games” and shows how the rich array of concepts from game theory may be combined with DEA to determine fair divisions of the rewards that are possible from multi-attribute efforts that are cooperatively employed to obtain these benefits in differing amounts. This chapter is also of interest from the standpoint of game theory because of its bearing on problems of computational implementability. Here the problems are given computationally implementable form.

Chapter 16 treats “stochastic frontier analysis (SFA)” — a regression based approach which is sometimes regarded as an alternative to DEA in evaluating performance efficiencies. SFA is also referred to as a “composed error regression” since it decomposes the usual error term into two components: (a) an “inefficiency component” and (b) a “random component” that represents things like (1) measurement error and (2) environmental influence such as the state of the economy in which a DMU is operating. As a result of recent work it is now possible to distinguish and separate the latter two items that were both previously embedded in the random error component of SFA. This makes further advances available for use in identifying different factors that are effecting performance.

In Chapter 16 this separation and identification is accomplished with a 3-stage process that combines DEA with SFA. A question may be raised concerning how this relates to the treatment of “non-discretionary variables” and their evaluation that is described in the earlier Chapter 7. The difference lies in the fact that this earlier treatment of non-discretionary variables is cross-sectional while the new 3 stage process described in Chapter 16 reveals time dependent differences in behavior. This property is illustrated by applying the 3-stage DEA-SFA approach of Chapter 16 to the Japanese banking industry. This is done for the period 1997-2001 which includes as environmental factors both the 1997/98 Asian financial crisis and the bursting of the IT bubble in 2001, both of which are beyond the control of management. A first stage analysis shows that the efficiency of these Japanese banks deteriorated over this period. However, filtering out the environmental factors shows that management efficiency increased substantially during this same period. Moreover, it did so consistently, as indicated by the trend of the SBM performance scores. Thus it would be unfair and probably counterproductive to penalize the managements of these banks for the overall deterioration of entity performances when, in fact, management had improved.

3. Conclusion and Acknowledgements

As already noted, these additional chapters are more advanced than the earlier ones. The earlier chapters, however, provide the background for these later chapters. As presented here these earlier chapters contain revisions and corrections from the first edition of this book, partly in response to reader comments and suggestions. We are grateful for these comments and hope that we can look forward to similar responses from readers of this second edition. We therefore reproduce the Preface to the first edition in the immediately following section of this book where, again, acknowledgement is due to Michiko Tone in Tokyo, Japan, and Bisheng Gu in Austin, Texas, for their efforts in typing and correcting the ms. in its various versions en route to completion.

WILLIAM W. COOPER

The Red McCombs School of Business
 University of Texas at Austin
 Austin, Texas 78712-1175, USA
 cooperw@mail.utexas.edu

LAWRENCE M. SEIFORD

Department of Industrial and
 Operations Engineering
 University of Michigan
 1205 Beal Avenue Ann Arbor
 MI 48109-2117, USA
 seiford@umich.edu

KAORU TONE

National Graduate Institute for Policy Studies
 7-22-1 Roppongi, Minato-ku
 Tokyo 106-8677, Japan
 tone@grips.ac.jp

June, 2006

W.W. Cooper wants to express his appreciation to the IC2 Institute of The University of Texas at Austin for continuing support of his research. Kaoru Tone would like to express his sincere gratitude to his colleagues and students: Bires K. Sahoo, Ken Nakabayashi, Miki Tsutsui and Junming Liu for their help in writing this Second Edition.

Notes

1. This term refers to the entities that are regarded as responsible for converting inputs into outputs.

Preface to the First Edition

1. Introduction

Recent years have seen a great variety of applications of DEA (Data Envelopment Analysis) for use in evaluating the performances of many different kinds of entities engaged in many different activities in many different contexts in many different countries. One reason is that DEA has opened up possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in many of these activities (which are often reported in non-commeasurable units). Examples include the maintenance activities of U.S. Air Force bases in different geographic locations, or police forces in England and Wales as well as performances of branch banks in Cyprus and Canada and the efficiency of universities in performing their education and research functions in the U.S., England and France. These kinds of applications extend to evaluating the performances of cities, regions and countries with many different kinds of inputs and outputs that include "social" and "safety-net" expenditures as inputs and various "quality-of-life" dimensions as outputs which, in turn, have led to dealing with important issues such as identifying sites for new locations (away from Tokyo) for the capital of Japan.

DEA has also been used to supply new insights into activities (and entities) that have previously been evaluated by other methods. For instance, studies of benchmarking practices with DEA have identified numerous sources of inefficiency in some of the most profitable firms — firms that served as benchmarks by reference to their (profitability only) criterion. DEA studies of the efficiency of different legal-organization forms as in "stock" vs. "mutual" insurance companies have shown that previous studies have fallen short in their attempts to evaluate the potentials of these different forms of organizations. (See below.) Similarly, a use of DEA has suggested reconsideration of previous studies of the efficiency with which pre- and post-merger activities of banks have been conducted.

The study of insurance company organization forms referenced above can be used to show not only how new results can be secured but also to show some of the new methods of data exploitation that DEA makes available. In order to study the efficiency of these organization forms it is, of course, necessary to remove other sources of inefficiency from the observations — unless one wants to *assume* that all such inefficiencies are absent.

In the study referenced in Chapter 7 of this book, this removal is accomplished by first developing separate efficiency frontiers for each of the two forms — stock vs. mutual. Each firm is then brought onto its respective frontier by removing its inefficiencies. A third frontier is then erected by reference to the thus adjusted data for each firm with deviations from this overall frontier then representing organization inefficiency. Statistical tests using the Mann-Whitney rank order statistic are then used in a manner that takes into account the full array of firms in each of the two categories. This led to the conclusion that

stock companies were uniformly more efficient in all of the dimensions that were studied.

In Chapter 7 this study is contrasted with the earlier studies by E. Fama and M. Jensen.¹ Using elements from “agency theory” to study the relative efficiency of these two forms of organization, Fama and Jensen concluded that each type of organization was relatively most efficient in supplying its special brand of services.

Lacking access to methods like those described above, Fama and Jensen utilized agency theory constructs which, like other parts of micro-economics, assume that performances of all firms occur on efficiency frontiers. This assumption can now be called into question by the hundreds of studies in DEA and related approaches² which have shown it to be seriously deficient for use with empirical data. Moreover, it is no longer necessary to make this assumption since the requisite conditions for fully efficient performances can be fulfilled by using concepts and procedures like those we have just described — and which are given implementable form in the chapters that follows.³

Fama and Jensen also base their conclusions on relatively few summary ratios which they reenforce by observing the long standing co-existence of mutual and stock companies. Left unattended, however, is what Cummins and Zi (1998, p.132)⁴ refer to as a “wave of conversions of insurers from mutual to stock ownership which is accelerating as competition increases from non-traditional services such as banks, mutual funds and security banks.” See also the discussion in Chapter 7 which notes that this movement has generally taken the form of “demutualizations” with no movement in the opposite direction.

In contrast to the relatively few summary ratios used by Fama and Jensen, DEA supplies a wealth of information in the form of estimates of inefficiencies in both inputs and outputs for every DMU (= Decision Making Unit). It also identifies the peer (comparison) group of efficient firms (stock and mutual) used to obtain these estimates and effect these evaluations. Moreover, the two-stage analysis described above provides additional information. Stage two identifies inefficiencies in each input and output which are attributable to the organization form used. Stage 1 identifies inefficiencies in each input and output attributable to the way each of these organization forms is managed. (P.L. Brockett and B. Golany (1996)⁵ suggest a use of the MDI (Minimum Discrimination Information) Statistic to determine whether crossovers occur with the possibility that one form of organization may be more efficient in some areas and less efficient in others — and hence automatically supplies a test of the Fama and Jensen hypothesis in a way that identifies where the crossovers, with accompanying advantages, occur.)

Turning to methodology, we emphasize that the linear (and mathematical) programming models and methods used in DEA effect their evaluations from observed (i.e., already executed) performances and hence reverse the usual manner in which programming models are used. This has led to new results as well as new insights for use in other parts of the programming literature. This use of linear programming has also led to new principles of inference from empirical

data which are directed to obtaining best estimates for *each* of the observed entities (or activities) in a collection of entities. This is in contrast to the usual *averaging* over all of the observations which are the characteristic approaches used in statistics and accounting. These estimates thus take the form of identifying the sources and amounts of inefficiency in each input and each output for every entity while also providing an overall measure of efficiency for *each* entity or activity that may be of interest. An illustrative example of the importance of this difference is made evident in a study of North Carolina hospitals each of which exhibited increasing or decreasing returns to scale (by DEA) whereas a use of “translog regression” they showed only constant returns to scale to prevail in “all.” See R.D. Banker, R.F. Conrad and R.P. Strauss (1986).⁶

2. Motivation

The great amount of activity involved in these uses of DEA have been accompanied by research directed to expanding its methods and concepts is evidenced by the hundreds of publications which are referenced in this book and elsewhere in the literature. The references, which appear in the attached disk (at the end of this text) are comprehensive but not exhaustive. Indeed, they are directed only to published materials and do not even attempt to cover the still larger (and increasing) numbers of unpublished studies and reports which are now in circulation which range from Ph.D. theses to brief notes and summaries. “A Bibliography of Data Envelopment Analysis (1978-2001)” by G. Tavares (GTavares@RUTCOR.Rutgers.edu) references more than 3600 papers, books, etc., by more than 1600 authors in 42 countries that include evaluating police forces in Australia and the joint economic and environmental performances of public utilities in Europe as well as the social progress of Arab societies in the Middle East and North Africa. Additional references may be secured at the Business School of the University of Warwick in England: <http://www.csv.warwick.ac.uk/~bsrlu/>. See also The Productivity Analysis Research Network (PARN), Odense University, Denmark: PARN@SAM.SDU.DK. In addition the developments in DEA have been (and are being) reported in two different literatures: (1) the literature of Operations Research/Management Science and (2) the literature of Economics.

This text is directed to an audience of users as well as students and the concepts are developed accordingly. For instance, efficiency considerations which are central to the DEA evaluations of interest are introduced by using the familiar and very simple ratio definition of “output divided by input.” This ratio formulation is then extended to multiple outputs and multiple inputs in a manner that makes contact with concepts embedded in more complex formulations. This includes the Pareto concept of “welfare efficiency” used in economics which is here referred to as “Pareto-Koopmans efficiency” in recognition of the adaptation of this “welfare economics” concept by T.C. Koopmans for use in “production economics.”

3. Methodology

The methodology used in this text is based mainly on linear algebra (including matrices and vectors) rather than the less familiar “set theory” methods used in some of the other texts. An advantage of this approach is that it very naturally and easily makes contact with the linear programming methods and concepts on which the developments in this book rest.

The power of this programming approach is greatly enhanced, as is well known, by virtue of mathematical duality relations which linear programming provides access to in unusually simple forms. This, in turn, provides opportunities for extending results and simplifying proofs which are not available from approaches that have tended to slight these linear-programming-duality relations. One reason for such slighting may be attributed to a lack of familiarity (or appreciation) of these concepts by the audiences to which these other texts were directed. To comprehend this possibility we have included an appendix — Appendix A — which provides a summary of linear programming and its duality relations and extends this to more advanced concepts such as the strong form of the “complementary slackness” principle.

Proofs are supplied for results which are critical but not obvious. Readers not interested in proof details can, if they wish, take the statements in the theorems on faith or belief. They can also check results and perhaps firm up their understanding by working the many small numerical examples which are included to illustrate theorems and their potential uses. In addition many examples are supplied in the form of miniaturized numerical illustrations.

To facilitate the use of this material a “DEA-Solver” disk is supplied with accompanying instructions which are introduced early in the text. In addition more comprehensive instructions for use of “DEA-Solver” are provided in Appendix B. This Solver was developed using VBA (Visual Basic for Applications) and Excel Macros in Microsoft Office 97 (a trademark of Microsoft Corporation) and is completely compatible with Excel data sheets. It can read a data set directly from an Excel worksheet and returns the results of computation to an Excel workbook. The code works on Excel 2000 as well. The results provide both primal (envelopment form) and dual (multiplier form) solutions as well as slacks, projections onto efficient frontiers, graphs, etc. The linear programming code was originally designed and implemented for DEA, taking advantage of special features of DEA formulations. Readers are encouraged to use this Solver on examples and problems in this text and for deepening their understanding of the models and their solutions. Although the attached Solver is primarily for learning purposes and can deal only with relatively small sized problems within a limited number of models, a more advanced “Professional version” can be found at the web site <http://www.saitech-inc.com/>.

Numerous specific references are also supplied for those who want to follow up some of the developments and example applications provided in this text. Still further references may be secured from the bibliography appended to the disk.

4. Strategy of Presentation

To serve the wider audience that is our objective, we have supplied numerous problems and suggested answers at the end of each chapter. The latter are referred to as “suggested responses” in order to indicate that other answers are possible in many cases and interested readers are encouraged to examine such possibilities.

In order to serve practitioner as well as classroom purposes we have presented these suggested responses immediately after the problems to which they refer. We have also provided appendices in cases where a deeper understanding or a broadened view of possibilities is desirable. We have thus tried to make this volume relatively self contained.

A wealth of developments in theory and methodology, computer codes, etc., have been (and are being) stimulated by the widespread use of DEA on many different problems in many different contexts. It is not possible to cover them all in a single text so, as already noted, we supply references for interested readers. Most of these references, including the more advanced texts that are also referenced, will require some knowledge of DEA. Supplying this requisite background is an objective of this text.

The first 5 chapters should suffice for persons seeking an introduction to DEA. In addition to different DEA models and their uses, these chapters cover different kinds of efficiency — from “technical” to “mix” or “allocative” and “returns-to-scale” inefficiencies. These topics are addressed in a manner that reflects the “minimal assumption” approach of DEA in that information (= data) requirements are also minimal. Additional information such as a knowledge of unit prices and costs or other weights such as weights stated in “utils” is not needed. Treatments of these topics are undertaken in Chapter 8 which deals with allocation (= allocative) efficiency on the assumption that access to the requisite information in the form of an exact knowledge of prices, costs, etc., is available.

Relaxation of the need for such “exact” information can be replaced with bounds on the values of the variables to be used. This topic is addressed in Chapter 6 because it has wider applicability than its potential for use in treating allocative inefficiency.

Chapter 7 treats variables for which the values cannot be completely controlled by users. Referred to as being “non-discretionary,” such variables are illustrated by weather conditions that affect the “sortie rates” that are reported by different air force bases or the unemployment rates that affect the recruitment potentials in different offices (and regions) of the U.S. Army Recruiting Command.⁷

Categorical (= classificatory) variables are also treated in Chapter 7. As exemplified by outlets which do or do not have drive-through capabilities in a chain of fast-food outlets, these variables are also non-discretionary but need to be treated in a different manner. Also, as might be expected, different degrees of discretion may need to be considered so extensions to the standard treatments of these topics are also included in Chapter 7.

It is characteristic of DEA that further progress has been made in treating data which are imperfectly known. Such imperfect knowledge may take a form in which the data can only be treated in the form of ordinal relations such as “less” or “more than.” See, for example, Ali, Cook and Seiford (1991).⁸ It can also take the form of knowing only that the values of the data lie within limits prescribed by certain upper and lower bounds. See Cooper, Park and Yu (1999)⁹ which unifies all of these approaches. See also Yu, Wei and Brockett (1996)¹⁰ for an approach to unification of DEA models and methods in a manner which relates them to multiple objective programming. Nor is this all, as witness recent articles which join DEA to “multiple objective programming” and “fuzzy sets.” See Cooper (2005)¹¹ and K. Triantis and O. Girod (1998).¹²

We again emphasize that the optimizations in DEA are directed to securing best estimates of performance for *each* observation (and the entities associated with them). In the numerous simulation studies that have now been performed this orientation has consistently led to results that favor DEA over alternate approaches that are directed optimizing over *all* observations (e.g., as in a least square regressions). See Banker, Chang and Cooper¹³ and the references cited therein.

This does not end the possibilities. It is also possible to combine DEA with statistical methods. As described in Chapter 9 this is done in the following two stage manner: In stage one DEA is applied to distinguish between efficient and inefficient observations of performances. In a succeeding stage two these results are incorporated as “dummy variable” in an ordinary least squares regression. Again the results, as reported in Bardham, Cooper and Kumbhakar (1998),¹⁴ showed that this approach gave better results than not only the ordinary least squares regression but also the stochastic frontiers regression that had been specifically developed to identify and evaluate inefficient performances in these same observations. See Kumbhakar and Lovell (2000)¹⁵ for a detailed development of these types of regression. As is apparent, much is going on with DEA in both research and uses. This kind of activity is likely to increase. Indeed, we hope that this volume will help to stimulate further such developments. The last chapter — *viz.*, Chapter 9 — is therefore presented in this spirit of continuing research in joining DEA with other approaches. These include modern techniques of “bootstrapping” as well as the classical “maximum likelihood” methods.

There is, of course, more to be done and the discussions in Chapter 9 point up some of these possibilities for research. This is the same spirit in which the other topics are treated in Chapter 9.

Another direction for research is provided by modifying the full (100%) efficiency evaluations of DEA by invoking the “satisficing” concepts of H.A. Simon,¹⁶ as derived from the literatures of management and psychology. Using “chance-constrained programming” approaches, a DEA model is presented in Chapter 9 which replaces the usual deterministic requirements of 100% efficiency as a basis for evaluation in DEA with a less demanding objective of achieving a “satisfactory” level of efficiency with a sufficiently high probability. This formulation

is then expanded to consider possible behaviors when a manager aspires to one level and superior echelons of management prescribe other (higher) levels, including levels which are probabilistically impossible of achievement.

Sensitivity of the DEA evaluations to data variations are topics that are also treated in Chapter 9. Methods for effecting such studies range from studies that consider the ranges over which a single data point may be varied without altering the efficiency ratings and extend to methods for examining effects on evaluations when all data are varied simultaneously — e.g., by worsening the output-input data for efficient performers and improving the output-input data for inefficient performers.

Such sensitivity analyses are not restricted to varying the components of given observations. Sensitivity analyses extend to eliminating some observations entirely and/or introducing additional observations. This topic is studied in Chapter 9 by reference to a technique referred to as “window analysis.” Originally viewed as a method for studying trends, window analysis is reminiscent of the use of “moving averages” in statistical time series. Here, however, it is also treated as a method for studying the stability of DEA results because such window analyses involve the removal of entire sets of observations and their replacement by other (previously not considered) observations.

Comments and suggestions for further research are offered as each topic is covered in Chapter 9. These comments identify shortcomings that need to be addressed as well as extensions that might be made. The idea is to provide readers with possible directions and “head starts” in pursuing any of these topics. Other uses are also kept in mind, however, as possibilities for present use are covered along with openings for further advances.

5. Acknowledgments

The authors of this text have all been actively involved in applications and uses of DEA that could not have been undertaken except as part of a team effort. We hope that the citations made in this text will serve as acknowledgments to the other team members who helped to make these applications possible, and who also helped to point up the subsequent advances in DEA that are described in this text.

Acknowledgment is also due to Michiko Tone in Tokyo, Japan, and Bisheng Gu in Austin, Texas. In great measure this text could not have been completed without their extraordinary efforts in typing, correcting and compiling the many versions this manuscript passed through en route to its present form. All of the authors benefited from their numerous suggestions as well as their efforts in coordinating these activities which spanned an ocean and a continent.

Finally, thanks are also due to Gary Folven of Kluwer Publishers who encouraged, cajoled and importuned the authors in manners that a lawyer might say, “aided and abetted them in the commission of this text.”

Mistakes that may be detected should be attributed only to the authors and we will be grateful if they are pointed out to us.

WILLIAM W. COOPER

The McCombs School of Business
University of Texas at Austin
Austin, Texas 78712-1175, USA
cooperw@mail.utexas.edu

LAWRENCE M. SEIFORD

Mechanical and Industrial Engineering
University of Massachusetts-Amherst
Amherst, MA 01003-2210
seiford@ecs.umass.edu

KAORU TONE

National Graduate Institute for Policy Studies
2-2 Wakamatsu-cho, Shinjuku-ku
Tokyo 162-8677, Japan
tone@grips.ac.jp

June, 1999

Kaoru Tone would like to express his heartfelt thanks to his colleagues and students: A. Hoodith, K. Tsuboi, S. Iwasaki, K. Igarashi and H. Shimatsuji for their help in writing this book. W.W. Cooper wants to express his appreciation to the IC² Institute of the University of Texas at Austin for their continuing support of his research. Larry Seiford would like to acknowledge the many DEA researchers worldwide whose collective efforts provide the backdrop for this book, and he especially wants to thank K. Tone and W.W. Cooper for their dedicated efforts in producing this text.

Notes

1. E.F. Fama and M.C. Jensen (1983), "Separation of Ownership and Control," *Journal of Law and Economics* 26, pp.301-325. See also, in this same issue, Fama and Jensen "Agency Problems and Residual Claims," pp.327-349.
2. E.g., the stochastic frontier regression approaches for which descriptions and references are supplied in Chapter 9.
3. See the discussion following (3.22) and (3.33) in Chapter 3 - which extend to identifying short vs. long run efficiencies in performance evaluation without recourse to time series.
4. J.D. Cummins and H. Zi (1998) "Comparisons of Frontier Efficiency Methods: An Application to the U.S. Life Insurance Industry," *Journal of Productivity Analysis* 10, pp.131-152.
5. P.L. Brockett and B. Golany (1996) "Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis," *Management Science* 42, pp.466-472.
6. R.D. Banker, R.F. Conrad and R.P. Strauss (1986), "A Comparative Application of Data Envelopment Analysis and Translog Methods: An Illustrative Study of Hospital Production," *Management Science* 32, pp.30-44.
7. See D.A. Thomas (1990) *Data Envelopment Analysis Methods in the Management of Personnel Recruitment under Competition in the Context of U.S. Army Recruiting*, Ph.D. Thesis (Austin, Texas: The University of Texas Graduate School of Business.) Also available from University Micro-Films, Inc., Ann Arbor, Michigan. Also see M.J. Kwon (2000) *Evaluating Military Recruitment to Determine the Relative Efficiencies of Joint vs. Service-Specific Advertising*, Ph.D. Thesis (Austin, Texas: The Red McCombs School of Business.)
8. A.I. Ali, W.D. Cook and L.M. Seiford (1991) "Strict vs Weak Ordinal Relations for Multipliers in Data Envelopment Analysis," *Management Science* 37, pp.733-738.
9. W.W. Cooper, K.S. Park and G. Yu (1999) "IDEA and AR: IDEA: Models for Dealing with Imprecise Data in DEA," *Management Science* 45, pp.597-607.
10. G. Yu, Q. Wei and P. Brockett (1996) "A Generalized Data Envelopment Model: A Unification and Extension Method for Efficiency Analysis of Decision Making Units," *Annals of Operations Research* 66, pp.47-92.
11. W.W. Cooper "Origins, Uses of and Relation Between Goal Programming and Multicriteria Programming," *Journal of Multi Criteria Decision Analysis* (to appear, 2005).
12. K. Triantis and O. Girod (1998) "A Mathematical Programming Approach for Measuring Technical Efficiency in a Fuzzy Environment," *Journal of Productivity Analysis* 10, pp.85-102.
13. R.D. Banker, H. Chang and W.W. Cooper (2004), "A Simulation Study of DEA and Parametric Frontier Models in the Presence of Heteroscedasticity," *European Journal of Operational Research* 153, pp.624-640.
14. I. Bardhan, W.W. Cooper and S. Kumbhakar (1998), " A Simulation Study of Joint Use of DEA and Stochastic Regressions for Production Function Estimation and Efficiency Evaluations," *Journal of Productivity Analysis* 9, pp.249-278.
15. S. Kumbhakar and C.A.K. Lovell (2000), *Stochastic Frontier Analysis* (Cambridge University Press).
16. H.A. Simon (1957) *Models of Man: Social and Rational* (New York: John Wiley & Sons, Inc.). See also his autobiography, H.A. Simon (1991) *Models of My Life* (New York: Basic Books).

Glossary of symbols

$\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}$	Small letters in bold face denote vectors.
$\mathbf{x} \in R^n$	\mathbf{x} is a point in the n dimensional vector space R^n .
$A \in R^{m \times n}$	A is a matrix with m rows and n columns.
\mathbf{x}^T, A^T	The symbol T denotes transposition.
e	e denotes a row vector in which all elements are equal to 1.
e_j	e_j is the unit row vector with the j -th element 1 and others 0.
I	I is the identity matrix.
$\text{rank}(A)$	$\text{rank}(A)$ denotes the rank of the matrix A .
$\sum_{j=1}^n \lambda_j$	indicates the sum of $\lambda_1 + \lambda_2 + \dots + \lambda_n$.
$\sum_{j=1, j \neq k}^n \lambda_j$	indicates the sum with λ_k omitted: $\lambda_1 + \dots + \lambda_{k-1} + \lambda_{k+1} + \dots + \lambda_n$.

1 GENERAL DISCUSSION

1.1 INTRODUCTION

This book is concerned with evaluations of performance and it is especially concerned with evaluating the activities of organizations such as business firms, government agencies, hospitals, educational institutions, etc. Such evaluations take a variety of forms in customary analyses. Examples include cost per unit, profit per unit, satisfaction per unit, and so on, which are measures stated in the form of a ratio like the following,

$$\frac{\text{Output}}{\text{Input}}. \quad (1.1)$$

This is a commonly used measure of efficiency. The usual measure of “productivity” also assumes a ratio form when used to evaluate worker or employee performance. “Output per worker hour” or “output per worker employed” are examples with sales, profit or other measures of output appearing in the numerator. Such measures are sometimes referred to as “partial productivity measures.” This terminology is intended to distinguish them from “total factor productivity measures,” because the latter attempt to obtain an output-to-input ratio value which takes account of *all* outputs and *all* inputs. Moving from partial to total factor productivity measures by combining all inputs and all outputs to obtain a single ratio helps to avoid imputing gains to one factor (or one output) that are really attributable to some other input (or output). For instance, a gain in output resulting from an increase in capital or improved

management might be mistakenly attributed to labor (when a single output to input ratio is used) even though the performance of labor *deteriorated* during the period being considered. However, an attempt to move from partial to total factor productivity measures encounters difficulties such as choosing the inputs and outputs to be considered and the weights to be used in order to obtain a single-output-to-single-input ratio that reduces to a form like expression (1.1).

Other problems and limitations are also incurred in traditional attempts to evaluate productivity or efficiency when multiple outputs and multiple inputs need to be taken into account. Some of the problems that need to be addressed will be described as we proceed to deal in more detail with Data Envelopment Analysis (DEA), the topic of this book. The relatively new approach embodied in DEA does not require the user to prescribe weights to be attached to each input and output, as in the usual index number approaches, and it also does not require prescribing the functional forms that are needed in statistical regression approaches to these topics.

DEA utilizes techniques such as mathematical programming which can handle large numbers of variables and relations (constraints) and this relaxes the requirements that are often encountered when one is limited to choosing only a few inputs and outputs because the techniques employed will otherwise encounter difficulties. Relaxing conditions on the number of candidates to be used in calculating the desired evaluation measures makes it easier to deal with complex problems and to deal with other considerations that are likely to be confronted in many managerial and social policy contexts. Moreover, the extensive body of theory and methodology available from mathematical programming can be brought to bear in guiding analyses and interpretations. It can also be brought to bear in effecting computations because much of what is needed has already been developed and adapted for use in many prior applications of DEA. Much of this is now available in the literature on research in DEA and a lot of this has now been incorporated in commercially available computer codes that have been developed for use with DEA. This, too, is drawn upon in the present book and a CD with supporting DEA-Solver software and instructions, has been included to provide a start by applying it to some problems given in this book.

DEA provides a number of additional opportunities for use. This includes opportunities for collaboration between analysts and decision-makers, which extend from collaboration in choices of the inputs and outputs to be used and includes choosing the types of “what-if” questions to be addressed. Such collaborations extend to “benchmarking” of “what-if” behaviors of competitors and include identifying potential (new) competitors that may emerge for consideration in some of the scenarios that might be generated.

1.2 SINGLE INPUT AND SINGLE OUTPUT

To provide a start to our study of DEA and its uses, we return to the single output to single input case and apply formula (1.1) to the following simple

example. Suppose there are 8 branch stores which we label A to H at the head of each column in Table 1.1.

Table 1.1. Single Input and Single Output Case

Store	A	B	C	D	E	F	G	H
Employee	2	3	3	4	5	5	6	8
Sale	1	3	2	3	4	2	3	5
Sale/Employee	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

The number of employees and sales (measured in 100,000 dollars) are as recorded in each column. The bottom line of Table 1.1 shows the sales per employee — a measure of “productivity” often used in management and investment analysis. As noted in the sentence following expression (1.1), this may also be treated in the more general context of “efficiency.” Then, by this measure, we may identify B as the most efficient branch and F as least efficient.

Let us represent these data as in Figure 1.1 by plotting “number of employees” on the horizontal and “sales” on the vertical axis. The slope of the line connecting each point to the origin corresponds to the sales per employee and the highest such slope is attained by the line from the origin through B . This line is called the “efficient frontier.” Notice that this frontier touches at least one point and all points are therefore on or below this line. The name Data Envelopment Analysis, as used in DEA, comes from this property because in mathematical parlance, such a frontier is said to “envelop” these points.

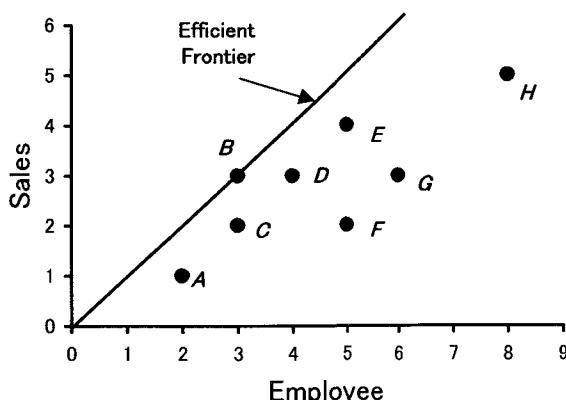


Figure 1.1. Comparisons of Branch Stores

Given these data, one might be tempted to draw a statistical regression line fitted to them. The dotted line in Figure 1.2 shows the regression line passing through the origin which, under the least squares principle, is expressed by $y = 0.622x$. This line, as normally determined in statistics, goes through the “middle” of these data points and so we could define the points above it as *excellent* and the points below it as *inferior* or *unsatisfactory*. One can measure the degree of excellence or inferiority of these data points by the magnitude of the deviation from the thus fitted line. On the other hand, the frontier line designates the performance of the best store (*B*) and measures the efficiency of other stores by deviations from it. There thus exists a fundamental difference between statistical approaches via regression analysis and DEA. The former reflects “average” or “central tendency” behavior of the observations while the latter deals with best performance and evaluates all performances by deviations from the frontier line. These two points of view can result in major differences when used as methods of evaluation. They can also result in different approaches to improvement. DEA identifies a point like *B* for future examination or to serve as a “benchmark” to use in seeking improvements. The statistical approach, on the other hand, averages *B* along with the other observations, including *F* as a basis for suggesting where improvements might be sought.

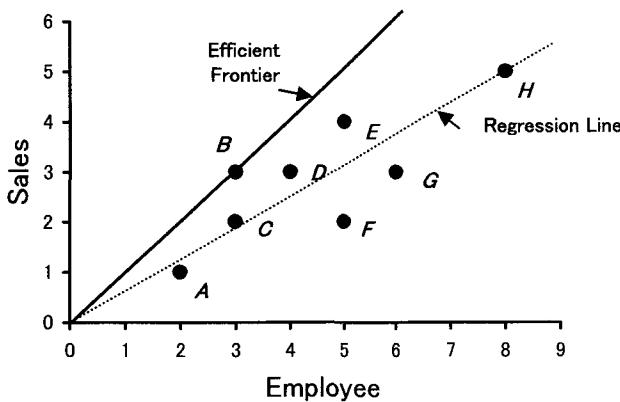


Figure 1.2. Regression Line vs. Frontier Line

Returning to the example above, it is not really reasonable to believe that the frontier line stretches to infinity with the same slope. We will analyze this problem later by using different DEA models. However, we assume that this line is effective in the range of interest and call it the *constant returns-to-scale* assumption.

Compared with the best store *B*, the others are inefficient. We can measure the efficiency of others relative to *B* by

$$0 \leq \frac{\text{Sales per employee of others}}{\text{Sales per employee of } B} \leq 1 \quad (1.2)$$

and arrange them in the following order by reference to the results shown in Table 1.2.

$$1 = B > E > D > C > H > A = G > F = 0.4.$$

Thus, the worst, F , attains $0.4 \times 100\% = 40\%$ of B 's efficiency.

Table 1.2. Efficiency

Store	A	B	C	D	E	F	G	H
Efficiency	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

Now we observe the problem of how to make the inefficient stores efficient, i.e., how to move them up to the efficient frontier. For example, store A in Figure 1.3 can be improved in several ways. One is achieved by reducing the input (number of employees) to A_1 with coordinates $(1, 1)$ on the efficient frontier. Another is achieved by raising the output (sales in \$100,000 units) up to $A_2(2, 2)$. Any point on the line segment $\overline{A_1A_2}$ offers a chance to effect the *improvements* in a manner which assumes that the input should not be increased and the output should not be decreased in making the store efficient.



Figure 1.3. Improvement of Store A

This very simple example moves from the ratio in Table 1.1 to the “ratio of ratios” in Table 1.2, which brings to the fore an important point. The values in (1.1) depend on the units of measure used whereas this is not the case for (1.2). For instance, if sales were stated in units of \$10,000, the ratio for F would change from $2/5 = 0.4$ to $20/5 = 4.0$. However, the value of (1.2) would remain unchanged at $4/10 = 0.4$ and the *relative efficiency* score associated with F is not affected by this choice of a different unit of measure. This property, sometimes referred to as “units invariance” has long been recognized as important in

engineering and science. Witness, for instance, the following example from the field of combustion engineering where ratings of furnace efficiency are obtained from the following formula,¹

$$0 \leq E_r = \frac{y_r}{y_R} \leq 1 \quad (1.3)$$

where

y_r = Heat obtained from a given unit of fuel by the furnace being rated,

y_R = Maximum heat that can be obtained from this same fuel input.

The latter, i.e., the maximum heat can be calculated from thermodynamic principles by means of suitable chemical-physical analyses. The point to be emphasized here, however, is that x , the amount of fuel used must be the same so that, mathematically,

$$0 \leq \frac{y_r/x}{y_R/x} = \frac{y_r}{y_R} \leq 1 \quad (1.4)$$

Hence, (1.3) is obtained from a ratio of ratios that is “units invariant.”

Returning to the ratios in Table 1.2, we might observe that these values are also bounded by zero and unity. However, the variations recorded in Table 1.2 may result from an excess amount of input or a deficiency in the output. Moreover, this situation is general in the business and social-policy (economics) applications which are of concern in this book. This is one reason we can make little use of formulas like (1.4). Furthermore, this formula is restricted to the case of a single output and input. Attempts to extend it to multiple inputs and multiple outputs encounter the troubles which were identified in our earlier discussion of “partial” and “total factor productivity” measures.

1.3 TWO INPUTS AND ONE OUTPUT CASE

To move to multiple inputs and outputs and their treatment, we turn to Table 1.3 which lists the performance of 9 supermarkets each with two inputs and one output. *Input* x_1 is the number of employees (unit: 10), *Input* x_2 the floor area (unit: $1000m^2$) and *Output* y the sales (unit: 100,000 dollars). However, notice that the sales are unitized to 1 under the constant returns-to-scale assumption. Hence, input values are normalized to values for getting 1 unit of sales. We plot the stores, taking *Input* x_1 /*Output* y and *Input* x_2 /*Output* y as axes which we may think of as “unitized axes” in Figure 1.4.

Table 1.3. Two Inputs and One Output Case

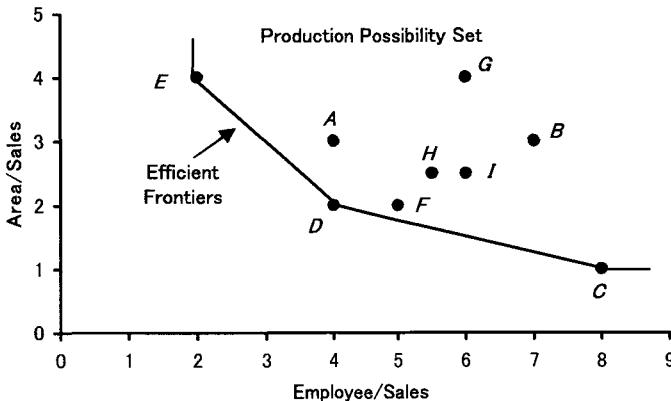


Figure 1.4. Two Inputs and One Output Case

From the efficiency point of view, it is natural to judge stores which use less inputs to get one unit output as more efficient. We therefore identify the line connecting C , D and E as the efficient frontier. We do not discuss the tradeoffs between these three stores but simply note here that no point on this frontier line can improve one of its input values without worsening the other. We can envelop all the data points within the region enclosed by the frontier line, the horizontal line passing through C and the vertical line through E . We call this region the *production possibility set*. (More accurately, it should be called the *piecewise linear* production possibility set assumption, since it is not guaranteed that the (true) boundary of this region is piecewise linear, i.e., formed of linear segments like the segment connecting E and D and the segment connecting D and C .) This means that the observed points are assumed to provide (empirical) evidence that production is possible at the rates specified by the coordinates of any point in this region.

The efficiency of stores not on the frontier line can be measured by referring to the frontier point as follows. For example, A is inefficient. To measure its inefficiency let \overline{OA} , the line from zero to A , cross the frontier line at P (see Figure 1.5). Then, the efficiency of A can be evaluated by

$$\frac{OP}{OA} = 0.8571.$$

This means that the inefficiency of A is to be evaluated by a combination of D and E because the point P is on the line connecting these two points. D and E are called the *reference set* for A . The reference set for an inefficient store may differ from store to store. For example, B has the reference set composed of C and D in Figure 1.4. We can also see that many stores come together around D and hence it can be said that D is an efficient store which is also “representative,” while C and E are also efficient but also possess unique characteristics in their association with segments of the frontiers that are far removed from any observations.

Now we extend the analysis in Figure 1.3 to identify improvements by referring inefficient behaviors to the efficient frontier in this two inputs (and one output) case. For example, A can be effectively improved by movement to P with *Input* $x_1 = 3.4$ and *Input* $x_2 = 2.6$, because these are the coordinates of P , the point on the efficient frontier that we previously identified with the line segment \overline{OA} in Figure 1.5. However, any point on the line segment $\overline{DA_1}$ may also be used as a candidate for improvement. D is attained by reducing *Input* x_2 (floor area), while A_1 is achieved by reducing *Input* x_1 (employees). Yet another possibility for improvement remains by increasing output and keeping the *status quo* for inputs. This will be discussed later.



Figure 1.5. Improvement of Store A

1.4 ONE INPUT AND TWO OUTPUTS CASE

Table 1.4 shows the number of customers (unit=10) per salesman and the sales (unit=100,000 dollars) per salesman of 7 branch offices. To obtain a unitized frontier in this case, we divide by the number of employees (=salesmen) which is considered to be the only input of interest. The efficient frontier then consists of the lines connecting B , E , F and G as shown in Figure 1.6.

Table 1.4. One Input and Two Outputs Case

Store	A	B	C	D	E	F	G
Employees	x	1	1	1	1	1	1
Customers	y_1	1	2	3	4	4	5
Sales	y_2	5	7	4	3	6	2

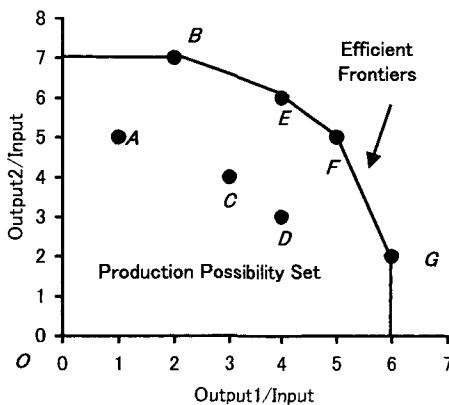


Figure 1.6. One Input and Two Outputs Case

The production possibility set is the region bounded by the axes and the frontier line. Branches A , C and D are inefficient and their efficiency can be evaluated by referring to the frontier lines. For example, from Figure 1.7, the efficiency of D is evaluated by

$$\frac{d(O, D)}{d(O, P)} = 0.75, \quad (1.5)$$

where $d(O, D)$ and $d(O, P)$ mean “distance from zero to D ” and “distance from zero to P ,” respectively.

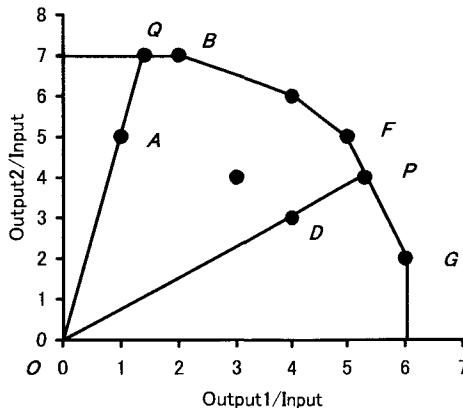


Figure 1.7. Improvement

The above ratio is referred to as a “radial measure” and can be interpreted as the ratio of two distance measures. The choice of distance measures is not

unique so,² because of familiarity, we select the Euclidean measures given by

$$d(O, D) = \sqrt{4^2 + 3^2} = 5$$

$$d(O, P) = \sqrt{\left(\frac{16}{3}\right)^2 + 4^2} = \frac{20}{3},$$

where the terms under the radical sign are squares of the coordinates of D and P , respectively, as obtained from Table 1.4 for D and from the intersection of $y_2 = \frac{3}{4}y_1$ and $y_2 = 20 - 3y_1$ for P . As claimed, substitution in (1.5) then gives

$$5 \div \frac{20}{3} = \frac{15}{20} = 0.75.$$

This interpretation as a ratio of distances aligns the results with our preceding discussion of such ratios. Because the ratio is formed relative to the Euclidean distance from the origin over the production possibility set, we will always obtain a measure between zero and unity.

We can also interpret the results for managerial (or other) uses in a relatively straightforward manner. The value of the ratio in (1.5) will always have a value between zero and unity. Because we are concerned with output, however, it is easier to interpret (1.5) in terms of its reciprocal

$$\frac{d(O, P)}{d(O, D)} = \frac{20}{3} \div 5 = 1.33.$$

This result means that, to be efficient, D would have had to increase both of its outputs by $4/3$. To confirm that this is so we simply apply this ratio to the coordinates of D and obtain

$$\frac{4}{3}(4, 3) = \left(\frac{16}{3}, 4\right),$$

which would bring coincidence with the coordinates of P , the point on the efficient frontier used to evaluate D .

Returning to (1.5) we note that 0.75 refers to the proportion of the output that P shows was possible of achievement. It is important to note that this refers to the proportion of inefficiency present in *both* outputs by D . Thus, the shortfall in D 's output can be repaired by increasing both outputs without changing their proportions — until P is attained.

As might be expected, this is only one of the various types of inefficiency that will be of concern in this book. This kind of inefficiency which can be eliminated without changing proportions is referred to as “technical inefficiency.”

Another type of inefficiency occurs when only some (but not all) outputs (or inputs) are identified as exhibiting inefficient behavior. This kind of inefficiency is referred to as “mix inefficiency” because its elimination will alter the proportions in which outputs are produced (or inputs are utilized).³

We illustrated the case of “technical inefficiency” by using D and P in Figure 1.7. We can use Q and B to illustrate “mix inefficiency” or we can use A , Q and

B to illustrate both technical and mix inefficiency. Thus, using the latter case we identify the technical efficiency component in A 's performance by means of the following radial measure,

$$\frac{d(O, A)}{d(O, Q)} = 0.714. \quad (1.6)$$

Using the reciprocal of this measure, as follows, and applying it to the coordinates of A at $(1, 5)$ gives

$$\frac{1}{0.714}(1, 5) = (1.4, 7),$$

as the coordinates of Q .

We can now note that the thus adjusted outputs are in the ratio $1.4/7=1/5$, which is the same as the ratio for A in Table 1.4 — viz., $y_1/y_2 = 1/5$. This augments both of the outputs of A without worsening its input and without altering the output proportions. This improvement in technical efficiency by movement to Q does not remove all of the inefficiencies. Even though Q is on the frontier it is not on an efficient part of the frontier. Comparison of Q with B shows a shortfall in output 1 (number of customers served) so a further increase in this output can be achieved by a lateral movement from Q to B . Thus this improvement can also be achieved without worsening the other output or the value of the input. Correcting output value, y_1 , without altering y_2 will change their proportions, however, and so we can identify two sources of inefficiencies in the performance of A : first a technical inefficiency via the radial measure given in (1.6) followed by a mix inefficiency represented by the output shortfall that remains in y_1 after all of the technical inefficiencies are removed.

We now introduce the term “purely technical inefficiency” so that, in the interest of simplicity, we can use the term “technical inefficiency” to refer to all sources of waste — purely technical and mix — which can be eliminated without worsening any other input or output. This also has the advantage of conforming to usages that are now fixed in the literature. It will also simplify matters when we come to the discussion of prices, costs and other kinds of values or weights that may be assigned to the different sources of inefficiency.

Comment : The term “technical efficiency” is taken from the literature of economics where it is used to distinguish the “technological” aspects of production from other aspects, generally referred to as “economic efficiency” which are of interest to economists.⁴ The latter involves recourse to information on prices, costs or other value considerations which we shall cover later in this text. Here, and in the next two chapters, we shall focus on purely technical and mix inefficiencies which represent “waste” that can be justifiably eliminated without requiring additional data such as prices and costs. It only requires assuring that the resulting improvements are worthwhile even when we do not specifically assign them a value.

As used here, the term mix inefficiency is taken from the accounting literatures where it is also given other names such as “physical variance” or

“efficiency variance.”⁵ In this usage, the reference is to physical aspects of production which exceed a prescribed standard and hence represent excessive uses of labor, raw materials, etc.

1.5 FIXED AND VARIABLE WEIGHTS

The examples used to this point have been very limited in the number of inputs and outputs used. This made it possible to use simple graphic displays to clarify matters but, of course, this was at the expense of the realism needed to deal with the multiple inputs and multiple outputs that are commonly encountered in practice. The trick is to develop approaches that make it possible to deal with such applications without unduly burdening users with excessive analyses or computations and without requiring large numbers of (often arbitrary or questionable) assumptions.

Consider, for instance, the situation in Table 1.5 which records behavior intended to serve as a basis for evaluating the relative efficiency of 12 hospitals in terms of two inputs, number of doctors and number of nurses, and two outputs identified as number of outpatients and inpatients (each in units of 100 persons/month).

Table 1.5. Hospital Case

Hospital	A	B	C	D	E	F	G	H	I	J	K	L
Doctors	20	19	25	27	22	55	33	31	30	50	53	38
Nurses	151	131	160	168	158	255	235	206	244	268	306	284
Outpatients	100	150	160	180	94	230	220	152	190	250	260	250
Inpatients	90	50	55	72	66	90	88	80	100	100	147	120

One way to simplify matters would be to weight the various inputs and outputs by pre-selected (fixed) weights. The resulting ratio would then yield an index for evaluating efficiencies. For instance, the weight

$$v_1(\text{weight for doctor}) : v_2(\text{weight for nurse}) = 5 : 1$$

$$u_1(\text{weight for outpatient}) : u_2(\text{weight for inpatient}) = 1 : 3$$

would yield the results shown in the row labelled “Fixed” of Table 1.6. (Notice that these ratios are normalized so that the maximum becomes unity, i.e., by dividing by the ratio of A.) This simplifies matters for use, to be sure, but raises a host of other questions such as justifying the 5 to 1 ratio for doctor vs. nurse and/or the 3 to 1 ratio of the weights for inpatients and outpatients. Finally, and even more important, are problems that can arise with the results shown – since it is not clear how much of the efficiency ratings are due to the weights and how much inefficiency is associated with the observations.

Table 1.6. Comparisons of Fixed vs. Variable Weights

Hospital	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>
Fixed	1	.90	.77	.89	.74	.64	.82	.74	.84	.72	.83	.87
CCR	1	1	.88	1	.76	.84	.90	.80	.96	.87	.96	.96

DEA, by contrast, uses variable weights. In particular, the weights are derived directly from the data with the result that the numerous *a priori* assumptions and computations involved in fixed weight choices are avoided. Moreover, the weights are chosen in a manner that assigns a best set of weights to each hospital. The term “best” is used here to mean that the resulting input-to-output ratio for each hospital is maximized relative to all other hospitals when these weights are assigned to these inputs and outputs for every hospital. The row labelled CCR in Table 1.6 shows results obtained from DEA using what is called the “CCR model”⁶ in DEA. As can be seen, these efficiency values are always at least as great as the ratio values obtained from the previous fixed value weights. Moreover, this “best ratio” result is general, under the following conditions: (1) all data and all weights are positive (or at least nonnegative), (2) the resulting ratio must lie between zero and unity and (3) these same weights for the target entity (=hospital) are applied to all entities. Consequently, the entity being evaluated cannot choose a better set of weights for its evaluation (relative to the other entities). The meaning of these results is clear. In each case, the evaluation is effected from a point on the efficient frontier so that a value like .88 for hospital *C* means that it is 12% inefficient. That is, compared to members of an efficient reference set, it is possible to identify a purely technical inefficiency of 12%—and possible mix inefficiencies as well—even under the best set of weights that *each* of these hospitals could choose to evaluate its own inefficiencies.

As we shall later see, the sources of inefficiency, such as purely technical and mix inefficiencies are automatically identified for each entity by DEA and their amounts estimated. Moreover, the reference set used to benchmark these inefficiencies are also identified. Finally, as we shall also see, these results are obtained using only minimal *a priori* assumptions. In addition to avoiding a need for *a priori* choices of weights, DEA does not require specifying the form of the relation between inputs and outputs in, perhaps, an arbitrary manner and, even more important, it does not require these relations to be the same for each hospital.

1.6 SUMMARY AND CONCLUSION

We have now covered a variety of topics which will be refined and extended in this book. Employing commonly used output-to-input ratio measures we

related them to topics such as measurements of productivity as well as the efficiency evaluation methods commonly used in economics, business and engineering. We will subsequently introduce other (non-ratio) approaches but will do so in ways that maintain contact with these ratio forms.

Extensions to multiple outputs and multiple inputs were examined in terms of fixed weights to be applied uniformly to the inputs and outputs of all entities to be evaluated, as in economic indices of "total factor productivity." This usage was then contrasted with the use of variable weights based on a best set being chosen for *each* entity to be evaluated, as in DEA. We also described interpretations and uses of the latter as derived from the efficient frontiers from which the evaluations were effected. This was then contrasted with the mixed, generally unclear, sources of inefficiencies that are implicit in the use of fixed weight approaches.

Additional advantages of DEA were also noted in terms of (a) its ability to identify sources and amounts of inefficiency in each input and each output for each entity (hospital, store, furnace, etc.) and (b) its ability to identify the benchmark members of the efficient set used to effect these evaluations and identify these sources (and amounts) of inefficiency.

All of the thus assessed entities were assumed to use the same inputs to produce the same outputs. Also, all data were assumed to be positive and weight choices were also restricted to positive values. These assumptions will be maintained in the immediately following chapters and then relaxed. Inputs and outputs were also assumed to be variable at the discretion of managers or designers. This assumption will also be maintained and then relaxed so that we will be able to distinguish between discretionary and non-discretionary inputs and outputs — to allow for differences in the circumstances under which different entities operate. Then we shall also introduce categorical variables to allow for further difference such as rural vs. urban categories, etc., to obtain more refined evaluations and insights.

The discussion in this chapter was confined to physical aspects of efficiency with distinctions between "purely technical" and "mix" inefficiencies. These were referred to as "waste" because they could be removed from any input or output without worsening any other input and output. Other types of inefficiency covered in this book will involve movement along efficient frontiers and hence also involve exchanges or substitutions. Such movements may be effected to achieve returns-to-scale economies or to improve cost and profit performances. All such movements along frontiers, however, imply an absence of technical inefficiencies (=purely technical plus mix). Hence this topic will be the center of attention in the immediately following chapters.

1.7 PROBLEM SUPPLEMENT FOR CHAPTER 1

Problem 1.1

To deal with multiple inputs and outputs, a ratio like the following may be used.

$$\frac{\sum_{r=1}^s u_r y_r}{\sum_{i=1}^m v_i x_i} = \frac{u_1 y_1 + u_2 y_2 + \cdots + u_s y_s}{v_1 x_1 + v_2 x_2 + \cdots + v_m x_m}$$

where

- y_r = amount of output r
- u_r = weight assigned to output r
- x_i = amount of input i
- v_i = weight assigned to input i .

The weights may be (1) fixed in advance or (2) derived from the data. The former is sometimes referred as an *a priori* determination..

1. Assignment 1

The weights given in the text for use in Table 1.5 are as follows:

$$v_1 = 5, \quad v_2 = 1$$

$$u_1 = 1, \quad u_2 = 3.$$

Apply these results to the example of Table 1.5 and compare your answer to the first two rows of Table 1.6. Then, interpret your results.

Suggestion: Be sure to normalize all results by dividing with the ratio for A , which is $370/251 \doteq 1.474$. Notice that this division cancels all units of measure. This is not the same as “units invariance,” however, which means that a change in the unit of measure will not affect the solution, e.g., if the number of doctors were restated in units of “10 doctors” or any other unit, then resulting solution values would be the same if the solution is “units invariant.”

2. Assignment 2

The manager of Hospital B asks you to determine a set of weights which will improve its standing relative to A .

Suggested Answer:

- Using the data in Table 1.5 you could determine weights which bring B to a par with A by solving

$$\frac{100u_1 + 90u_2}{20v_1 + 151v_2} = \frac{150u_1 + 50u_2}{19v_1 + 131v_2}. \quad (1.7)$$

- An easier route would be to solve the following problem.

$$\max \frac{150u_1 + 50u_2}{19v_1 + 131v_2}$$

subject to

$$\frac{150u_1 + 50u_2}{19v_1 + 131v_2} \leq 1$$

$$\frac{100u_1 + 90u_2}{20v_1 + 151v_2} \leq 1.$$

with $u_1, u_2, v_1, v_2 > 0$. The choice of u_1, u_2 and v_1, v_2 should maximize the ratio for Hospital *B*, so no better choice can be made from this manager's standpoint.

- As shown in the next chapter, this nonlinear programming problem can be replaced by the following linear program.

$$\max 150u_1 + 50u_2$$

subject to

$$150u_1 + 50u_2 \leq 19v_1 + 131v_2$$

$$100u_1 + 90u_2 \leq 20v_1 + 151v_2$$

$$19v_1 + 131v_2 = 1$$

and all variables are constrained to be positive. Note that the normalization in the last constraint ensures that the weights will be relative. Since they sum to unity, no further normalization is needed such as the one used in the answer to Assignment 1. Also the possibility of zero values for all weights is eliminated, even though it is one possible solution for the equality pair (1.7) stated Assignment 2 for Problem 1.1.

Suggested Answer : Solutions to the first 2 problems involve treating nonlinear problems so we focus on this last problem. By the attached software DEA-Solver, we then obtain

$$u_1^* \doteq 0.00463, \quad u_2^* \doteq 0.00611, \quad v_1^* \doteq 0.0275, \quad v_2^* \doteq 0.00364,$$

where “ \doteq ” means “approximately equal to.” This solution is also optimal for the preceding problem with

$$\frac{150(0.00463) + 50(0.00611)}{19(0.0275) + 131(0.00364)} \doteq 1$$

for Hospital *B* and

$$\frac{100(0.00463) + 90(0.00611)}{20(0.0275) + 151(0.00364)} \doteq 0.921$$

for Hospital *A*. This reversal of efficiency ratings might lead to *A* responding by calculating his own best weights, and similarly for the other hospitals. A regulatory agency might then respond by going beyond such pairwise comparisons, in which case we could use an extension of the above approaches – to be described in succeeding chapters – which

effects such a “best rating” by considering all 12 hospitals simultaneously for each such evaluation. In fact this was the approach used to derive the evaluations in the row labelled “CCR” in Table 1.6. Note that Hospitals *A* and *B* are both rated as fully efficient when each is given its own “best” weights. With the exception of *D*, none of the other hospitals achieve full (=100%) DEA efficiency even when each is accorded its own “best” weights.

Suggestion for using the attached DEA-Solver

You can solve the best weights for the above two hospital problem using the supporting “DEA-Solver” software on the included CD. See Appendix B for installation instructions for a PC. Then follow the procedures below:

1. Create an Excel 97 file containing the data sheet as exhibited in Figure 1.8 where (I) and (O) indicate Input and Output, respectively.

	A	B	C	D	E	F
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4						

Figure 1.8. Excel File “HospitalAB.xls”

2. Save the file with the file name “HospitalAB.xls” in an appropriate folder and close Excel.
3. Start DEA-Solver and follow the instructions on the display.
4. Choose “CCR-I” as DEA model.
5. Choose “HospitalAB.xls” as data file.
6. Choose “Hospital.xls” as the Workbook for saving results of the computation.
7. Click “Run.”
8. After the computation is over, click the “Exit” button.
9. Open the sheet “Weight” which contains optimal weights obtained for each hospital. You will see results like in Table 1.7. From the table, we can see a set of optimal weights for Hospital *A* as given by

$$v_1^* = .025, v_2^* = 3.31E - 3 = 3.31 \times 10^{-3} = .00331$$

$$u_1^* = 3.74E - 3 = 3.74 \times 10^{-3} = .00374, u_2^* = 6.96E - 3 = 6.96 \times 10^{-3} = .00696$$

Table 1.7. Optimal Weights for Hospitals *A* and *B*

No.	DMU	Score	v(1)	v(2)	u(1)	u(2)
1	<i>A</i>	1	0.025	3.31E-03	3.74E-03	6.96E-03
2	<i>B</i>	1	2.75E-02	3.64E-03	4.63E-03	6.11E-03

and for Hospital *B* as

$$v_1^* = 2.75E-2 = 2.75 \times 10^{-2} = .0275, v_2^* = 3.64E-3 = 3.64 \times 10^{-3} = .00364$$

$$u_1^* = 4.63E-3 = 4.63 \times 10^{-3} = .00463, u_2^* = 6.11E-3 = 6.11 \times 10^{-3} = .00611.$$

These weights give the best ratio score 1 (100%) to each hospital.

However, notice that the best weights are not necessarily unique as you can see from the “Fixed” weight case in Table 1.6. Actually, the weights

$$v_1 : v_2 = 5 : 1, \quad u_1 : u_2 = 1 : 3$$

or more concretely,

$$v_1 = .02, \quad v_2 = .004, \quad u_1 = .0027, \quad u_2 = .0081$$

are applied for this fixed weight evaluation and these also made Hospital *A* efficient.

Problem 1.2

The comptroller and treasurer of an industrial enterprise discuss whether the company’s sales should be treated as an input or an output.

Suggested Resolution: Take the ratio of output ÷ input and ask whether an increase in sales should improve or worsen the company’s efficiency rating in terms of its effects on the value of this ratio. Compare this with whether you could treat an increase in expenses as an output or an input in terms of its effects on the ratio.

Problem 1.3

The ratios in Table 1.6 are designed for use in evaluating the performance efficiency of each hospital. This means that entire hospitals are to be treated as Decision Making Units (DMUs) which convert inputs of Doctors and Nurses into outputs of Inpatients and Outpatients. Can you suggest other DMUs to evaluate the performance of hospital?

Suggested Answer : David Sherman used surgical units to evaluate performances of teaching hospitals for the rate-setting Commission of Massachusetts

because outputs such as interns, residents and nurses to be trained in the performances of such services have proved difficult to treat with measures, such as cost/patient, which the Commission had been using.⁷

Problem 1.4

Suppose that a manager of a chain-store is trying to evaluate performance of each store. He selected factors for evaluation as follows: (1) the annual average salary per employee as input, (2) the number of employees as input, and (3) the annual average sales per employee as output. Criticize this selection.

Suggested Answer : Let p_i be the number of employees of store i , c_i be the total annual salary paid by store i , and d_i be the total annual sales of store i . Then the weighted ratio scale which expresses the manager's criterion would be

$$\frac{u_1 d_i / p_i}{v_1 c_i / p_i + v_2 p_i},$$

where u_1 = weight for output d_i / p_i (the average sales per employee), v_1 = weight for input c_i / p_i (the average salary per employee) and v_2 = weight for input p_i (the number of employees).

The above ratio can be transformed into

$$\frac{u_1 d_i}{v_1 c_i + v_2 p_i^2}.$$

This evaluation thus puts emphasis on the number of employees by squaring this value, while other factors are evaluated in a linear manner. If such uneven treatment has no special justification, it may be better to use a ratio form such as,

$$\frac{u_1 d_i}{v_1 c_i + v_2 p_i}.$$

This choice would be (1) the total salary paid by the store as input, (2) the number of employees as input, and (3) the total annual sales of the store as output.

You should be careful in dealing with processed data, e.g., value per head, average, percentile, and raw data at the same time.

Problem 1.5

Enumerate typical inputs and outputs for performance measurement of the following organizations: (1) airlines, (2) railways, (3) car manufacturers, (4) universities, (5) farms, and (6) baseball teams.

Notes

1. This example is taken from A. Charnes, W.W. Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, 1978, pp.429-444.

2. Our measures of distance and their uses are related to each other and discussed in W.W. Cooper, L.M. Seiford, K. Tone and J. Zhu "Some Models and Measures for Evaluating Performances with DEA: Past Accomplishments and Future Prospects," *Journal of Productivity Analysis* (forthcoming).

3. The latter is referred to as an input mix and the former as an output mix inefficiency.
4. See, for example, pp.15-18 in H.R. Varian *Microeconomic Analysis* 2nd ed. (New York. W.W. Norton & Co., 1984.)
5. Vide p.192 in W.W. Cooper and Y. Ijiri, eds., *Kohler's Dictionary For Accountants, 6th Edition* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1981.)
6. After Charnes, Cooper and Rhodes (1978) above.
7. See H.D. Sherman, "Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Other Techniques for Measuring and Locating Efficiency in Health Care Organizations," Ph.D. Thesis (Boston: Harvard University Graduate School of Business, 1981.) Also available from University Microfilms, Inc., Ann Arbor, Michigan.

2 THE BASIC CCR MODEL

2.1 INTRODUCTION

This chapter deals with one of the most basic DEA models, the *CCR model*, which was initially proposed by Charnes, Cooper and Rhodes in 1978. Tools and ideas commonly used in DEA are also introduced and the concepts developed in Chapter 1 are extended. There, for each DMU, we formed the virtual input and output by (yet unknown) weights (v_i) and (u_r):

$$\begin{aligned}\text{Virtual input} &= v_1x_{1o} + \cdots + v_mx_{mo} \\ \text{Virtual output} &= u_1y_{1o} + \cdots + u_sy_{so}.\end{aligned}$$

Then we tried to determine the weight, using linear programming so as to maximize the ratio

$$\frac{\text{virtual output}}{\text{virtual input}}.$$

The optimal weights may (and generally will) vary from one DMU to another DMU. Thus, the “weights” in DEA are derived from the data instead of being fixed in advance. Each DMU is assigned a best set of weights with values that may vary from one DMU to another. Additional details and the algorithms used to implement these concepts are explained in succeeding chapters.

2.2 DATA

In DEA, the organization under study is called a *DMU* (Decision Making Unit). The definition of DMU is rather loose to allow flexibility in its use over a wide range of possible applications. Generically a DMU is regarded as the entity responsible for converting inputs into outputs and whose performances are to be evaluated. In managerial applications, DMUs may include banks, department stores and supermarkets, and extend to car makers, hospitals, schools, public libraries and so forth. In engineering, DMUs may take such forms as airplanes or their components such as jet engines. For the purpose of securing *relative* comparisons, a group of DMUs is used to evaluate each other with each DMU having a certain degree of managerial freedom in decision making.

Suppose there are n DMUs: $DMU_1, DMU_2, \dots, DMU_n$. Some common input and output items for each of these $j = 1, \dots, n$ DMUs are selected as follows:

1. Numerical data are available for each input and output, with the data assumed to be positive¹ for all DMUs.
2. The items (inputs, outputs and choice of DMUs) should reflect an analyst's or a manager's interest in the components that will enter into the relative efficiency evaluations of the DMUs.
3. In principle, smaller input amounts are preferable and larger output amounts are preferable so the efficiency scores should reflect these principles.
4. The measurement units of the different inputs and outputs need not be congruent. Some may involve number of persons, or areas of floor space, money expended, etc.

Suppose m input items and s output items are selected with the properties noted in 1 and 2. Let the input and output data for DMU_j be $(x_{1j}, x_{2j}, \dots, x_{mj})$ and $(y_{1j}, y_{2j}, \dots, y_{sj})$, respectively. The input data matrix X and the output data matrix Y can be arranged as follows,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \quad (2.1)$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ y_{s1} & y_{s2} & \cdots & y_{sn} \end{pmatrix} \quad (2.2)$$

where X is an $(m \times n)$ matrix and Y an $(s \times n)$ matrix. For example, the hospital case in Section 1.5 has the data matrices:

$$X = \begin{pmatrix} 20 & 19 & 25 & 27 & 22 & 55 & 33 & 31 & 30 & 50 & 53 & 38 \\ 151 & 131 & 160 & 168 & 158 & 255 & 235 & 206 & 244 & 268 & 306 & 284 \end{pmatrix}$$

$$Y = \begin{pmatrix} 100 & 150 & 160 & 180 & 94 & 230 & 220 & 152 & 190 & 250 & 260 & 250 \\ 90 & 50 & 55 & 72 & 66 & 90 & 88 & 80 & 100 & 100 & 147 & 120 \end{pmatrix}$$

so x_{1j} = number of doctors and x_{2j} = number of nurses used by hospital j in servicing (= producing) y_{1j} = number of outpatients and y_{2j} = number of inpatients.

2.3 THE CCR MODEL

Given the data, we measure the efficiency of each DMU once and hence need n optimizations, one for each DMU_j to be evaluated. Let the DMU_j to be evaluated on any trial be designated as DMU_o where o ranges over 1, 2, ..., n . We solve the following fractional programming problem to obtain values for the input “weights” (v_i) ($i = 1, \dots, m$) and the output “weights” (u_r) ($r = 1, \dots, s$) as variables.

$$(FP_o) \quad \max_{\mathbf{v}, \mathbf{u}} \theta = \frac{u_1 y_{1o} + u_2 y_{2o} + \dots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \dots + v_m x_{mo}} \quad (2.3)$$

$$\text{subject to } \frac{u_1 y_{1j} + \dots + u_s y_{sj}}{v_1 x_{1j} + \dots + v_m x_{mj}} \leq 1 \quad (j = 1, \dots, n) \quad (2.4)$$

$$v_1, v_2, \dots, v_m \geq 0 \quad (2.5)$$

$$u_1, u_2, \dots, u_s \geq 0. \quad (2.6)$$

The constraints mean that the ratio of “virtual output” vs. “virtual input” should not exceed 1 for every DMU. The objective is to obtain weights (v_i) and (u_r) that maximize the ratio of DMU_o , the DMU being evaluated. By virtue of the constraints, the optimal objective value θ^* is at most 1. Mathematically, the nonnegativity constraint (2.5) is not sufficient for the fractional terms in (2.4) to have a positive value. We do not treat this assumption in explicit mathematical form at this time. Instead we put this in managerial terms by assuming that all outputs and inputs have some nonzero worth and this is to be reflected in the weights u_r and v_i being assigned some positive value.

2.4 FROM A FRACTIONAL TO A LINEAR PROGRAM

We now replace the above fractional program (FP_o) by the following linear program (LP_o),

$$(LP_o) \quad \max_{\boldsymbol{\mu}, \boldsymbol{v}} \theta = \mu_1 y_{1o} + \dots + \mu_s y_{so} \quad (2.7)$$

$$\text{subject to} \quad \nu_1 x_{1o} + \cdots + \nu_m x_{mo} = 1 \quad (2.8)$$

$$\mu_1 y_{1j} + \cdots + \mu_s y_{sj} \leq \nu_1 x_{1j} + \cdots + \nu_m x_{mj} \quad (2.9)$$

$$(j = 1, \dots, n)$$

$$\nu_1, \nu_2, \dots, \nu_m \geq 0 \quad (2.10)$$

$$\mu_1, \mu_2, \dots, \mu_s \geq 0. \quad (2.11)$$

Theorem 2.1 *The fractional program (FP_o) is equivalent to (LP_o) .*

Proof. Under the nonzero assumption of v and $X > 0$, the denominator of the constraint of (FP_o) is positive for every j , and hence we obtain (2.3) by multiplying both sides of (2.4) by the denominator. Next, we note that a fractional number is invariant under multiplication of both numerator and denominator by the same nonzero number. After making this multiplication, we set the denominator of (2.3) equal to 1, move it to a constraint, as is done in (2.8), and maximize the numerator, resulting in (LP_o) . Let an optimal solution of (LP_o) be $(\nu = \nu^*, \mu = \mu^*)$ and the optimal objective value θ^* . The solution $(v = v^*, u = \mu^*)$ is also optimal for (FP_o) , since the above transformation is reversible under the assumptions above. (FP_o) and (LP_o) therefore have the same optimal objective value θ^* . \square

We also note that the measures of efficiency we have presented are “units invariant” — i.e., they are independent of the units of measurement used in the sense that multiplication of each input by a constant $\delta_i > 0$, $i = 1, \dots, m$, and each output by a constant $p_r > 0$, $r = 1, \dots, s$, does not change the obtained solution. Stated in precise form we have

Theorem 2.2 (Units Invariance Theorem) *The optimal values of $\max \theta = \theta^*$ in (2.3) and (2.7) are independent of the units in which the inputs and outputs are measured provided these units are the same for every DMU.*

Thus, one person can measure outputs in miles and inputs in gallons of gasoline and quarts of oil while another measures these same outputs and inputs in kilometers and liters. They will nevertheless obtain the same efficiency value from (2.3) or (2.7) when evaluating the same collection of automobiles, say. See Note 2 for proof.²

Before proceeding we note that (LP_o) can be solved by the simplex method of linear programming. The optimal solution can be more easily obtained by dealing with the dual side of (LP_o) , however, as will be explained in detail in Chapter 3.

In any case let us suppose we have an optimal solution of (LP_o) which we represent by (θ^*, v^*, u^*) ³ where v^* and u^* are values with constraints given in (2.10) and (2.11). We can then identify whether *CCR-efficiency* has been achieved as follows:

Definition 2.1 (CCR-Efficiency)

1. *DMU_o is CCR-efficient if $\theta^* = 1$ and there exists at least one optimal (v^*, u^*) , with $v^* > \mathbf{0}$ and $u^* > \mathbf{0}$.*
2. *Otherwise, DMU_o is CCR-inefficient.*

Thus, CCR-inefficiency means that either (i) $\theta^* < 1$ or (ii) $\theta^* = 1$ and at least one element of (v^*, u^*) is zero for every optimal solution of (LP_o) . We will explain the latter case using an example in Section 2.6.2, and a detailed description of CCR-efficiency will be given in Chapter 3.

Now we observe the case where DMU_o has $\theta^* < 1$ (CCR-inefficient). Then there must be at least one constraint (or DMU) in (2.9) for which the weight (v^*, u^*) produces equality between the left and right hand sides since, otherwise, θ^* could be enlarged. Let the set of such $j \in \{1, \dots, n\}$ be

$$E'_o = \{ j : \sum_{r=1}^s u_r^* y_{rj} = \sum_{i=1}^m v_i^* x_{ij} \}. \quad (2.12)$$

The subset E_o of E'_o , composed of CCR-efficient DMUs, is called the *reference set* or the *peer group* to the DMU_o . It is the existence of this collection of efficient DMUs that forces the DMU_o to be inefficient. The set spanned by E_o is called the *efficient frontier* of DMU_o .

2.5 MEANING OF OPTIMAL WEIGHTS

The (v^*, u^*) obtained as an optimal solution for (LP_o) results in a set of optimal weights for the DMU_o . The ratio scale is evaluated by :

$$\theta^* = \frac{\sum_{r=1}^s u_r^* y_{ro}}{\sum_{i=1}^m v_i^* x_{io}}. \quad (2.13)$$

From (2.8), the denominator is 1 and hence

$$\theta^* = \sum_{r=1}^s u_r^* y_{ro}. \quad (2.14)$$

As mentioned earlier, (v^*, u^*) are the set of most favorable weights for the DMU_o in the sense of maximizing the ratio scale. v_i^* is the optimal weight for the input item i and its magnitude expresses how highly the item is evaluated, relatively speaking. Similarly, u_r^* does the same for the output item r . Furthermore, if we examine each item $v_i^* x_{io}$ in the virtual input

$$\sum_{i=1}^m v_i^* x_{io} (= 1), \quad (2.15)$$

then we can see the relative importance of each item by reference to the value of each $v_i^* x_{io}$. The same situation holds for $u_r^* y_{ro}$ where the u_r^* provides a measure of the relative contribution of y_{ro} to the overall value of θ^* . These values not only show which items contribute to the evaluation of DMU_o , but also to what extent they do so.

2.6 EXPLANATORY EXAMPLES

We illustrate the use of the CCR model via the following small-scale examples.

2.6.1 Example 2.1 (1 Input and 1 Output Case)

Table 2.1 shows 8 DMUs with 1 input and 1 output. (The first example in Chapter 1).

Table 2.1. Example 2.1

DMU	A	B	C	D	E	F	G	H
Input	2	3	3	4	5	5	6	8
Output	1	3	2	3	4	2	3	5

We can evaluate the efficiency of DMU A, by solving the LP problem below:

$$\begin{aligned} < A > \max \quad & \theta = u \\ \text{subject to} \quad & 2v = 1 \\ & u \leq 2v \quad (A) \quad 3u \leq 3v \quad (B) \\ & 2u \leq 3v \quad (C) \quad 3u \leq 4v \quad (D) \\ & 4u \leq 5v \quad (E) \quad 2u \leq 5v \quad (F) \\ & 3u \leq 6v \quad (G) \quad 5u \leq 8v \quad (H) \end{aligned}$$

where all variables are constrained to be nonnegative.

The optimal solution, easily obtained by simple ratio calculations, is given by ($v^* = 0.5$, $u^* = 0.5$, $\theta^* = 0.5$). Thus, the CCR-efficiency of A is $\theta^* = u^* = 0.5$. The reference set for A is found to be $E_A = \{B\}$ by inserting $u^* = 0.5$ and $v^* = 0.5$, the best possible weights for DMU A, in each of the above constraints. Thus the performance of B is used to characterize A and rates it as inefficient even with the best weights that the data admit for A.

The efficiency of B can be similarly evaluated from the data in Table 2.1 by:

$$\begin{aligned} < B > \max \quad & \theta = 3u \\ \text{subject to} \quad & 3v = 1 \\ & u \leq 2v \quad (A) \quad 3u \leq 3v \quad (B) \\ & 2u \leq 3v \quad (C) \quad 3u \leq 4v \quad (D) \\ & 4u \leq 5v \quad (E) \quad 2u \leq 5v \quad (F) \\ & 3u \leq 6v \quad (G) \quad 5u \leq 8v \quad (H) \end{aligned}$$

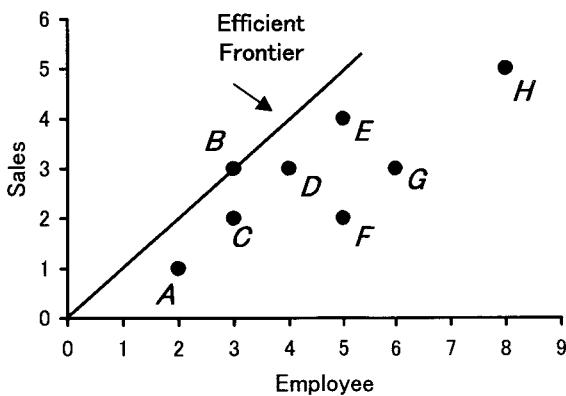
The optimal solution is ($v^* = 0.3333$, $u^* = 0.3333$, $\theta^* = 1$) and B is CCR-efficient. See Definition 2.1.

We can proceed in a similar way with the other DMUs to obtain the results shown in Table 2.2. Only DMU B is efficient and is in the reference set of all of the other DMUs. (See Figure 2.1.) Figure 2.1 portrays the situation geometrically. The efficient frontier represented by the solid line passes through B and no other point. The θ^* values in Table 2.2 show what is needed to bring each DMU onto the efficient frontier. For example, the value of $\theta^* = 1/2$ applied

Table 2.2. Results of Example 2.1

DMU	$CCR(\theta^*)$	Reference Set
A	0.5000	B
B	1.0000	B
C	0.6667	B
D	0.7500	B
E	0.8000	B
F	0.4000	B
G	0.5000	B
H	0.6250	B

to A 's input will bring A onto the efficient frontier by reducing its input 50% while leaving its output at its present value. Similarly $0.6667 \times 3 = 2$ will position C on the frontier. And so on.

**Figure 2.1.** Example 2.1

2.6.2 Example 2.2 (2 Inputs and 1 Output Case)

Table 2.3 shows 6 DMUs with 2 inputs and 1 output where the output value is unitized to 1 for each DMU.

(1) The linear program for DMU A is:

$$\begin{aligned}
 & \langle A \rangle \quad \max \quad \theta = u \\
 & \text{subject to} \quad 4v_1 + 3v_2 = 1 \\
 & \qquad \qquad u \leq 4v_1 + 3v_2 \quad (A) \quad u \leq 7v_1 + 3v_2 \quad (B) \\
 & \qquad \qquad u \leq 8v_1 + v_2 \quad (C) \quad u \leq 4v_1 + 2v_2 \quad (D) \\
 & \qquad \qquad u \leq 2v_1 + 4v_2 \quad (E) \quad u \leq 10v_1 + v_2 \quad (F)
 \end{aligned}$$

where all variables are constrained to be nonnegative.

Table 2.3. Example 2.2

	DMU	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
Input	x_1	4	7	8	4	2	10
	x_2	3	3	1	2	4	1
Output	y	1	1	1	1	1	1

This problem can be solved by a linear programming code. It can also be solved by simply deleting v_2 from the inequalities by inserting $v_2 = (1 - 4v_1)/3$ and observing the relationship between v_1 and u . The (unique) optimal solution is ($v_1^* = 0.1429$, $v_2^* = 0.1429$, $u^* = 0.8571$, $\theta^* = 0.8571$) and the CCR-efficiency of *A* is 0.8571. By applying the optimal solution to the above constraints, the reference set for *A* is found to be $E_A = \{D, E\}$.

(2) The linear program for DMU *B* is:

$$\begin{aligned} & \langle B \rangle \max \quad \theta = u \\ & \text{subject to} \quad 7v_1 + 3v_2 = 1 \\ & \qquad \qquad \qquad u \leq 4v_1 + 3v_2 \quad (A) \qquad u \leq 7v_1 + 3v_2 \quad (B) \\ & \qquad \qquad \qquad u \leq 8v_1 + v_2 \quad (C) \qquad u \leq 4v_1 + 2v_2 \quad (D) \\ & \qquad \qquad \qquad u \leq 2v_1 + 4v_2 \quad (E) \qquad u \leq 10v_1 + v_2 \quad (F) \end{aligned}$$

The (unique) optimal solution is ($v_1^* = 0.0526$, $v_2^* = 0.2105$, $u^* = 0.6316$, $\theta^* = 0.6316$), the CCR-efficiency of *B* is 0.6316, and the reference set is $E_B = \{C, D\}$.

Now let us observe the difference between the optimal weights $v_1^* = 0.0526$ and $v_2^* = 0.2105$. The ratio $v_2^*/v_1^* = 0.2105/0.0526 = 4$ suggests that it is advantageous for *B* to weight *Input* x_2 four times more than *Input* x_1 in order to maximize the ratio scale measured by virtual input vs. virtual output. These values have roles as measures of the sensitivity of efficiency scores in reference to variations in input items. This topic will be dealt with in detail in Chapter 9 of this book where a systematic basis for conducting such sensitivity analyses will be provided. Here we only note that our analysis shows that a reduction in *Input* x_2 has a bigger effect on efficiency than does a reduction in *Input* x_1 .

(3) An optimal solution for *C* is ($v_1^* = 0.0833$, $v_2^* = 0.3333$, $u^* = 1$, $\theta^* = 1$) and *C* is CCR-efficient by Definition 2.1. However, the optimal solution is not uniquely determined, as will be observed in the next section.

Likewise, *D* and *E* are CCR-efficient.

(4) The linear program for DMU *F* is:

$$\begin{aligned} & \langle F \rangle \max \quad \theta = u \\ & \text{subject to} \quad 10v_1 + v_2 = 1 \\ & \qquad \qquad \qquad u \leq 4v_1 + 3v_2 \quad (A) \qquad u \leq 7v_1 + 3v_2 \quad (B) \\ & \qquad \qquad \qquad u \leq 8v_1 + v_2 \quad (C) \qquad u \leq 4v_1 + 2v_2 \quad (D) \\ & \qquad \qquad \qquad u \leq 2v_1 + 4v_2 \quad (E) \qquad u \leq 10v_1 + v_2 \quad (F) \end{aligned}$$

The optimal solution for F is ($v_1^* = 0$, $v_2^* = 1$, $u^* = 1$, $\theta^* = 1$) and with $\theta^* = 1$, F looks efficient. However, we notice that $v_1^* = 0$. We therefore assign a small positive value ε to v_1 and observe the change in θ^* . That is, we use the data for F and with $10\varepsilon + v_2 = 1$ to obtain $v_2 = 1 - 10\varepsilon$. By inserting this value in the above inequalities, the following constraints are obtained:

$$\begin{array}{ll} u \leq 3 - 26\varepsilon & (A) \\ u \leq 1 - 2\varepsilon & (C) \\ u \leq 4 - 38\varepsilon & (E) \end{array} \quad \begin{array}{ll} u \leq 3 - 23\varepsilon & (B) \\ u \leq 2 - 16\varepsilon & (D) \\ u \leq 1 & (F) \end{array}$$

Noting that ε is a small positive value, the minimum of the right-hand terms is attained with

$$u = 1 - 2\varepsilon.$$

Therefore, for any $\varepsilon > 0$, it follows that $\theta^* = 1 - 2\varepsilon < 1$. Thus, v_1 must be zero in order for F to have $\theta^* = 1$. We therefore conclude that F is CCR-inefficient by Definition 2.1.

Furthermore, let us examine the inefficiency of F by comparing F with C . C has *Input* $x_1 = 8$ and *Input* $x_2 = 1$, while F has *Input* $x_1 = 10$ and *Input* $x_2 = 1$. F has 2 units of excess in *Input* x_1 compared with C . This deficiency is concealed because the optimal solution forces the weight of *Input* x_1 to zero ($v_1^* = 0$). C is in the reference set of F and hence by direct comparison we can identify the fact that F has used an excessive amount of this input.

It is not always easy to see such an excess in an input (or a shortage in output) from the optimal solution of the CCR model. In the next chapter, we will approach the CCR model from the dual side of the linear program and this will enable us to determine the excesses and shortfalls explicitly by the nonzero values with which these are identified.

A DMU such as F , with $\theta^* = 1$ and with an excess in inputs and/or a shortage in outputs, is called *ratio* efficient but *mix* inefficient.

Table 2.4 shows the CCR-efficiency (θ^*) of Example 2.2 and Figure 2.2 depicts the efficient frontier.

Table 2.4. Results of Example 2.2

DMU	x_1	x_2	y	CCR(θ^*)	Reference Set	v_1	v_2	u
A	4	3	1	0.8571	$D E$.1429	.1429	.8571
B	7	3	1	0.6316	$C D$.0526	.2105	.6316
C	8	1	1	1	C	.0833	.3333	1
D	4	2	1	1	D	.1667	.1667	1
E	2	4	1	1	E	.2143	.1429	1
F	10	1	1	1	C	0	1	1

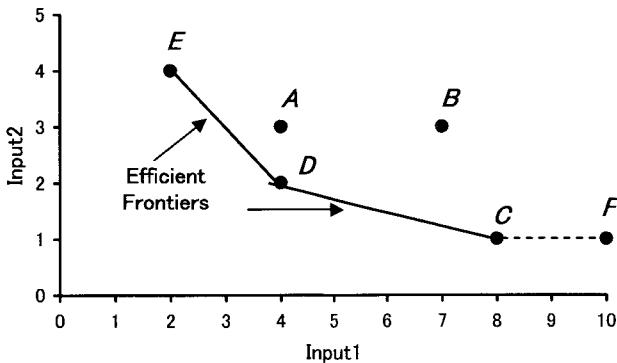


Figure 2.2. Example 2.2

2.7 ILLUSTRATION OF EXAMPLE 2.2

In order to demonstrate the role of weights (v, u) for identifying the CCR-efficiency of DMUs, we will show graphically the efficient frontier of Example 2.2 in the weight variables (=multiplier) space. Example 2.2 has 2 inputs and 1 output, whose value is unitized to 1. For this simple example we can illustrate the situations using a two dimensional graph. The linear programming constraints for each DMU have the following inequalities in common with all variables being constrained to be nonnegative.

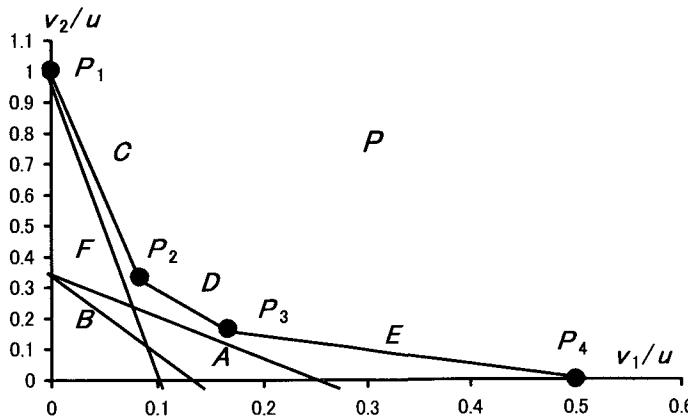
$$\begin{array}{ll} u \leq 4v_1 + 3v_2 & (A) \\ u \leq 8v_1 + v_2 & (C) \\ u \leq 2v_1 + 4v_2 & (E) \end{array} \quad \begin{array}{ll} u \leq 7v_1 + 3v_2 & (B) \\ u \leq 4v_1 + 2v_2 & (D) \\ u \leq 10v_1 + v_2 & (F) \end{array}$$

Dividing these expressions by $u > 0$, we obtain the following inequalities:

$$\begin{array}{ll} 1 \leq 4(v_1/u) + 3(v_2/u) & (A) \\ 1 \leq 8(v_1/u) + (v_2/u) & (C) \\ 1 \leq 2(v_1/u) + 4(v_2/u) & (E) \end{array} \quad \begin{array}{ll} 1 \leq 7(v_1/u) + 3(v_2/u) & (B) \\ 1 \leq 4(v_1/u) + 2(v_2/u) & (D) \\ 1 \leq 10(v_1/u) + (v_2/u) & (F) \end{array}$$

These inequalities are depicted in Figure 2.3 by taking v_1/u and v_2/u as axes. The area denoted by P then shows the feasible region for the above constraints. The boundary of P consists of three line segments and two axes. The three line segments correspond to the efficient DMUs C , D and E .

We explain this situation using D as an example and we also explain the relationship between this region and the inefficient DMUs using A as an example.

Figure 2.3. Region P

1. Example D

The linear program for D consists of the preceding inequalities plus the following,

$$\begin{aligned} & \max && u \\ & \text{subject to} && 4v_1 + 2v_2 = 1. \end{aligned} \quad (2.16)$$

Dividing (2.16) by u , we have

$$4(v_1/u) + 2(v_2/u) = 1/u. \quad (2.17)$$

The objective function $u \rightarrow \max$ yields the same solutions as $1/u \rightarrow \min$, so the problem is to find the minimum t for which the following line touches the region P :

$$4(v_1/u) + 2(v_2/u) = t. \quad (2.18)$$

From Figure 2.3 we see that $t = 1$ (and hence $u = 1$) represents the optimal line for D , showing that D is efficient. It is also easy to see that D is efficient for *any* weight (v_1, v_2) on the line segment (P_2, P_3) . This observation leads to the conclusion that the optimal (v_1, v_2) for D is not unique. In fact, the value $(v_1 = .1667, v_2 = .1667)$ for D in Table 2.4 is an example, and actually corresponds to P_3 in Figure 2.3. See Problem 2.1 at the end of this Chapter.

Similarly, any (v_1, v_2) on the line segment (P_1, P_2) expresses the optimal weight for C and any (v_1, v_2) on the line segment (P_3, P_4) for E .

Thus, the optimal weights for an efficient DMU need not be unique and we should be careful to keep this in mind.

2. Example A

Next, we consider the inefficient DMUs, taking A as an example. The linear

program for A consists of the following expressions added to the inequalities above:

$$\begin{aligned} & \max \quad u \\ & \text{subject to} \quad 4v_1 + 3v_2 = 1. \end{aligned} \quad (2.19)$$

As with example D , (2.19) can be transformed into

$$4(v_1/u) + 3(v_2/u) = t. \quad (2.20)$$

Then the problem is to find the minimum t within the region P . Referring to Figure 2.4, we can see that the solution is given by the point P_3 , where the line parallel to line A touches the region P for the first time. P_3 is the intersection of lines D and E and this is the geometric correspond of the fact that the reference set to A consists of D and E . A simple calculation finds that $t = 1/0.8571$ and hence the efficiency of A is $u = .8571$. The value of (v_1, v_2) at P_3 is :

$$v_1 = .1667 \times .8571 = .1429, \quad v_2 = .1667 \times .8571 = .1429, \quad (2.21)$$

which are the optimal weights for A . The optimal weights for A are unique. Usually, the optimal weights for inefficient DMUs are unique, the exception being when the line of the DMU is parallel to one of the boundaries of the region P .

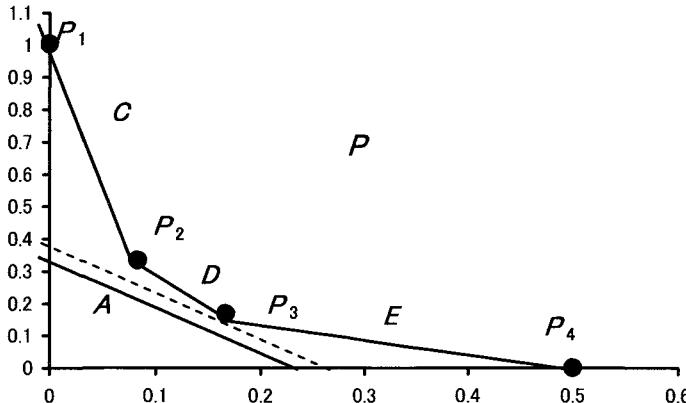


Figure 2.4. The Case of DMU A

2.8 SUMMARY OF CHAPTER 2

In this chapter, we introduced the CCR model, which is a basic DEA model.

- For each DMU, we formed the virtual input and output by (yet unknown) weights (v_i) and (u_r):

$$\text{Virtual input} = v_1x_{1o} + \cdots + v_mx_{mo}$$

$$\text{Virtual output} = u_1y_{1o} + \cdots + u_sy_{so}.$$

Then we tried to determine the weight, using linear programming so as to maximize the ratio

$$\frac{\text{virtual output}}{\text{virtual input}}.$$

The optimal weights may (and generally will) vary from one DMU to another DMU. Thus, the “weights” in DEA are derived from the data instead of being fixed in advance. Each DMU is assigned a best set of weights with values that may vary from one DMU to another. Here, too, the DEA weights differ from customary weightings (e.g., as in index number constructions) so we will hereafter generally use the term “multiplier” to distinguish these DEA values from the other commonly used approaches.

- CCR-efficiency was defined, along with the reference sets for inefficient DMUs.
- Details of the linear programming solution procedure and the production function correspondence are given in Chapter 3.

2.9 SELECTED BIBLIOGRAPHY

The term ‘Decision Making Unit’ (DMU) was used for the first time in the CCR model proposed in Charnes, Cooper and Rhodes (1978).⁴ The term DEA (Data Envelopment Analysis) was introduced in their report “A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiment in U.S. Public School Education,” (1978),⁵ Rhodes (1978)⁶ and appeared in Charnes, Cooper and Rhodes’ subsequent paper (1979).⁷ DEA originated from efforts to evaluate results from an early 1970’s project called “Program Follow Through”—a huge attempt by the U.S. Office (now Department) of Education to apply principles from the statistical design of experiments to a set of matched schools in a nationwide study. The purpose of the study was to evaluate educational programs designed to aid disadvantaged students in U.S. public schools. The data base was sufficiently large that issues of degrees of freedom, etc., were not a serious problem despite the numerous input and output variables used in the study. Nevertheless, unsatisfactory and even absurd results were secured from all of the statistical-econometric approaches that were tried. While trying to respond to this situation, Rhodes called Cooper’s attention to Farrell’s seminal article, “The Measurement of Productive Efficiency,” in the *Journal of the Royal Statistical Society* (1957). Charnes, Cooper and Rhodes extended Farrell’s work and succeeded in establishing DEA as a basis for efficiency analysis. Details of the project are described in Charnes, Cooper and Rhodes (1981).⁸ A brief history of DEA can be found in Charnes and Cooper (1985).⁹

2.10 PROBLEM SUPPLEMENT FOR CHAPTER 2

Problem 2.1

In Example 2.2, determine the region of (v_1, v_2) that makes each of DMUs C , D and E efficient, by referring to Table 2.4 and Figure 2.3.

Suggested Answer : For C : line segment $\overline{P_1P_2}$. This is the line segment stretching from $(v_1/u = 0, v_2/u = 1)$ at P_1 to $(v_1/u = .08333, v_2/u = .3333)$ at P_2 . For D : line segment $\overline{P_2P_3}$. This is the line segment stretching from $(v_1/u = .08333, v_2/u = .3333)$ at P_2 to $(v_1/u = .1667, v_2/u = .1667)$ at P_3 . For E : line segment $\overline{P_3P_4}$. This is the line segment stretching from $(v_1/u = .1667, v_2/u = .1667)$ at P_3 to $(v_1/u = .5, v_2/u = 0)$ at P_4 .

Problem 2.2

Use the data of Tables 2.3 and 2.4 to relate Figures 2.2 and 2.3.

Suggested Answer : The relation between Figures 2.2 and 2.3 is an example of what is called “the point-to-hyperplane correspondence” in mathematical programming. This means that the coordinates of the points for the representation in one of the spaces correspond to the coefficients for the hyperplanes in the dual (here=multiplier) space. The spaces in these figures are 2-dimensional, so the hyperplanes take the form of lines. For example, the coordinates for A in Figure 2.2 as obtained from Table 2.3 correspond to the coefficients in $4v_1 + 3v_2 \geq 1u$, the expression associated with DMU A as obtained from Table 2.4. In this single output case, we can use (2.19) to effect a further simplification by moving to homogenous coordinates and then try to minimize t , as given in (2.20). This minimization is undertaken subject to the similar transformation (to homogenous coordinates) for all of the other constraints obtained from Table 2.4. The result, as given in the discussion of (2.21), is $t = 1/0.8571$, so $u = 0.8571$ and $v_1 = v_2 = 0.1429$. Substitution in the expression with which we began then gives

$$1.00 \doteq 4v_1 + 3v_2 > 1u = 0.8571.$$

Hence DMU A is found to be inefficient. It should have produced more output, or used less input (or both).

This evaluation, as noted in Table 2.4, is determined relative to the corresponding expressions obtained for D and E — both of which achieve equality between both sides of their expressions, while using these same (best) weights for A — viz.

$$D : 0.8574 = 4v_1 + 2v_2 \approx 0.8571$$

$$E : 0.8574 = 2v_1 + 4v_2 \approx 0.8571.$$

We now note that the points for A and B in Figure 2.2 lie *above* the efficient frontier whenever the corresponding hyperplanes (here=lines) lie *below* the hyperplanes for D and E in Figure 2.3. The situation for F , which is also inefficient, differs because its hyperplane (=line) intersects the efficiency

frontier at one point, P_1 , but otherwise lies everywhere below it. This is the dual reflection of the hyperplane-to-point correspondence while examination of Figure 2.2 shows that F has the same coordinate as C for Input 2 but F 's coordinate for Input 1 exceeds the value of the same coordinate for C .

In this chapter we examined the multiplier model in which the maximizing objective tries to move each hyperplane “up,” as far as the other constraints allow, with the highest hyperplanes used to determine the boundaries of P , as exhibited in Figure 2.3. In the next chapter we shall examine the minimizing model which moves in the opposite direction in order to move the boundaries of P (the production possibility set) as far “down” as possible. Then we shall use the duality theorem of linear programming to show that the optimal values yield the same production possibility sets so either (or both) problems may be used, as desired and, even better, a suitable reading of the solution yields the solution to both problems when either one is solved.

Problem 2.3

Can you relate the CCR ratio form given by (2.3)-(2.6) in this chapter to the “ratio of ratios” form given by (1.4) in Chapter 1 for engineering efficiency?

Suggested Response : The “ratio of ratios” form given for engineering efficiency in (1.4) is for use with a single input and a single output. This can be interpreted as “virtual inputs” and “virtual outputs” for the values used in the numerators and denominators of (2.3)-(2.6) when specific values are assigned to all of the variables u_r and v_i . To put these ratios in the form of a “ratio of ratios” we will use what is called the TDT (Thompson-Dharmapala-Thrall) measure of efficiency obtained from the following problem,

$$\max_{\mathbf{u}, \mathbf{v}} \quad \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \Bigg/ \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}} \quad (2.22)$$

$$\text{where } \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}} = \text{maximum}_{j=1,\dots,n} \left\{ \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \right\} \quad (2.23)$$

$$u_r, v_i \geq 0 \quad \forall r, i. \quad (2.24)$$

Here “ $\forall r, i$ ” means “for all r, i .” Thus, for each choice of u_r and v_i , the maximum of the ratios in the braces on the right in the last expression is to be used as the denominator ratio in the first expression. Considering all allowable possibilities the problem is to maximize the ratio in (2.22) and “theoretical” as well as “observed” values may be used in these expressions if desired. A detailed discussion of the TDT measure and its properties may be found in W.W. Cooper, R.G. Thompson and R.M. Thrall “Extensions and New Development in DEA,” *Annals of Operations Research*, 66, 1996, pp.3-45. Here we only note that no bounds are placed on the admissible values in the ratios of the last expression. The idea is to allow these to be specialized, if one wants to do so, by imposing such bounds on any or all these ratios. If we limit all the ratio values to a bound of unity, as is done in (2.4), we obtain the model given

in (2.3)-(2.6). The value of the objective, (2.3), can then be interpreted as a “ratio of ratios” because a necessary condition for a solution to be optimal in (2.3)-(2.6) is that at least one of the ratios must achieve the maximal allowable value of unity. Thus, the maximization for (2.3)-(2.6) obtains an optimal value in the form of “ratio of ratios” in which the maximal term in the braces gives

$$\frac{\sum_{r=1}^s u_r^* y_{rk}}{\sum_{i=1}^m v_i^* x_{ik}} = 1.$$

Problem 2.4

Background: “Evaluating Efficiency of Turbofan Jet Engines in Multiple Input-Output Contexts: A Data Envelopment Analysis Approach,” by S. Bulla, W.W. Cooper, K.S. Park and D. Wilson(1999)¹⁰ reports results from a study of 29 jet engines produced by different manufacturers. The engineering measure of efficiency is given by $\eta = TV/\dot{Q}$ where T = Thrust of Engine, V = Cruise Velocity and \dot{Q} = Heat Input from Fuel. The DEA evaluations were based on these same two outputs but the single input (in the denominator) was replaced by 3 inputs: (1) Fuel Consumption (2) Weight of Engine and (3) Drag.

Using data from commercially available sources the results obtained from these two approaches to efficiency are portrayed in Figure 2.5 on the left. Coefficients for the resulting regression were found to be statistically significant. The value of R^2 , however, suggested that the explanatory power of engineering efficiency was low. More input variables are needed to bring the engineering results into closer correspondence with the DEA efficiency scores.

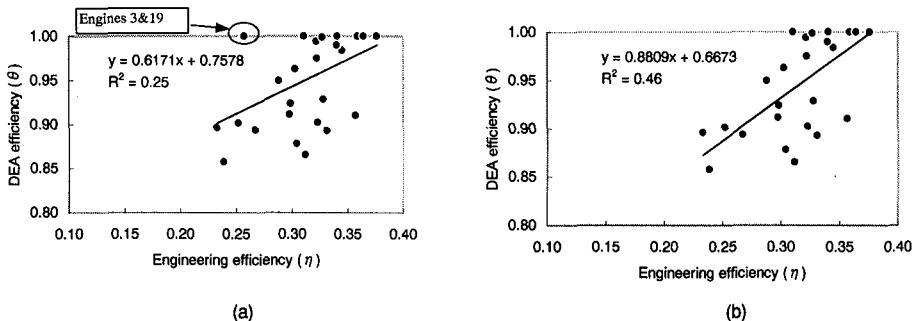


Figure 2.5. A Scatter Plot and Regression of DEA and Engineering Efficiency Ratings

As shown at the top of Figure (a), Engines 3 and 19 have the same coordinates. Their “outlier” character suggests that their removal would yield a regression with a higher R^2 . This is confirmed by the new regression and the new R^2 in Figure (b) where the estimates show statistically significant improve-

ments. Nevertheless, the same conclusion is reached: more inputs are needed for the engineering ratio to provide better “explanations” of the DEA scores, but it is not clear how this can be done.

Assignment : Is the removal of engines 3 and 19 justified or should the potential effects on the efficiency scores of other engines also be considered? Discuss.

Suggested Answer : Statistical independence for all observations is generally assumed in the removal of outliers. However, engines 3 and 19 have DEA ratings of unity and this suggests that they may be members of reference sets used to evaluate *other* engines. To check this possibility we use Figure 2.6 which is known as an “envelopment map.”¹¹

Eng.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	N*
1		✓													✓														✓	3
2		✓														✓													✓	3
3		✓																												0
4		✓														✓		✓											✓	4
5																✓		✓												2
6																	✓												✓	2
7																	✓	✓	✓											3
8																	✓	✓	✓									✓	4	
9																	✓	✓	✓									✓	4	
10																	✓	✓	✓									✓	4	
11																	✓												0	
12																		✓											✓	2
13																		✓												1
14																		✓												0
15																		✓												0
16																		✓												0
17																		✓	✓	✓								✓	4	
18																			✓										0	
19																			✓										0	
20																		✓	✓										2	
21																		✓										✓	2	
22																		✓										✓	2	
23		✓																✓		✓								✓	4	
24																		✓	✓	✓									3	
25																		✓	✓	✓									3	
26																		✓	✓	✓									3	
27																		✓	✓	✓									3	
28																		✓	✓	✓									3	
29																													✓	0
TN*	0	0	4	0	0	0	0	0	0	0	1	0	0	0	7	10	15	0	12	0	0	0	0	0	0	0	0	12	61	

TN* ≡ Total number of times that engine j ($= 1, \dots, 29$ in column) was used to evaluate *other* engines.

N* ≡ Number of times that *other* engines were used to evaluate engine i ($= 1, \dots, 29$ in row).

Figure 2.6. Envelopment Map for 29 Jet Engines

The values at the bottom of each column show the number of times an engine identified at the top of this column entered into a reference set to evaluate *other* engines, and the total for each row shows the number of *other* engines used in an evaluation. Thus, engine 3 was used a total (net) of 4 times in evaluating *other* engines and the row total shows that no *other* engine was involved in the evaluation of engine 3 — a property of efficient engines when the latter are extreme points of the production possibility set. A removal of engine 3 will therefore affect the efficiency ratings of other engines — unless in each case there exists an alternate optimum in which engine 3 does *not* actively enter as a member of the basis (=reference set).

No similar interaction with other engines occurs for engine 19, however, as is clear from the fact that the row and column totals are both zero for this engine. Hence our analysis shows that 19 (but not 3) may be treated as an outlier *without affecting any other observations*. Proceeding in this manner produces Figure 2.7 which, though statistically significant, is closer to Figure 2.5 (a).

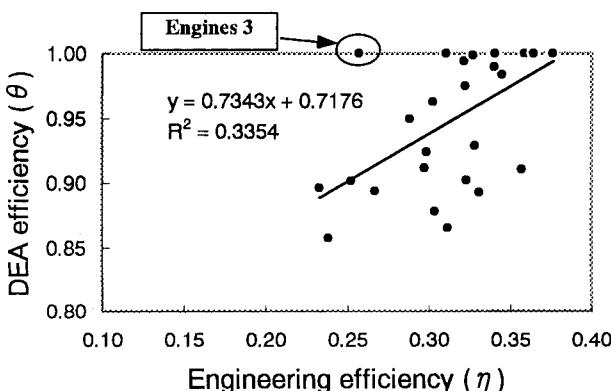


Figure 2.7. A Comparative Portrayal and Regression Analysis for 28 Engines (All Data Except for Engine 19)

Problem 2.5

Try to bring the engineering definition into play in a DEA evaluation.

Suggested Response : As noted in (1.4) the engineering “ratio of ratios” reduces to a comparison of actual to theoretically obtainable output from the amount of input actually used by any engine. This could be brought into play in a DEA evaluation by associating the theoretically obtainable output with the actual input for each engine (=DMU). This would produce n additional observations giving $2n$ DMUs from which evaluations would be made. However, there is no guarantee that a DMU will be evaluated relative to its own input. A DMU

evaluation of any engine (=DMU) could utilize a reference set yielding a *lower* input and a *higher* output. This could be checked and further analyzed, however, because the members of the reference set would be identified as well as the output shortfalls and input excesses for the actual DMU being evaluated.

Notes

1. This condition will be relaxed to allow nonnegative data in Chapter 3. Furthermore, in Chapter 5 (Section 5.2), we will introduce models which can also deal with negative data.

2. *Proof of Theorem 2.2.* Let θ^*, u_r^*, v_i^* be optimal for (2.3)-(2.6). Now replace the original y_{rj} and x_{ij} by $\rho_r y_{rj}$ and $\delta_i x_{ij}$ for some choices of $\rho_r, \delta_i > 0$. But then choosing $u'_r = u_r^*/\rho_r$ and $v'_i = v_i^*/\delta_i$ we have a solution to the transformed problem with $\theta' = \theta^*$. An optimal value for the transformed problem must therefore have $\theta'^* \geq \theta^*$. Now suppose we could have $\theta'^* > \theta^*$. Then, however, $u_r = u_r^* \rho_r$ and $v_i = v_i^* \delta_i$ satisfy the original constraints so the assumption $\theta'^* > \theta^*$ contradicts the optimality assumed for θ^* under these constraints. The only remaining possibility is $\theta'^* = \theta^*$. This proves the invariance claimed for (2.3). Theorem 2.1 demonstrated the equivalence of (LP_o) to (FP_o) and thus the same result must hold and the theorem is therefore proved. \square

3. We use the notations ν and μ instead of ν and μ in (LP_o) .

4. A. Charnes, W.W. Cooper and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, pp.429-444.

5. A. Charnes, W.W. Cooper and E. Rhodes (1978), "A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiments in U.S. Public School Education," Management Science Research Report No. 432, Carnegie-Mellon University, School of Urban and Public Affairs, Pittsburgh, PA.

6. E.L. Rhodes (1978), "Data Envelopment Analysis and Related Approaches for Measuring the Efficiency of Decision-Making Units with an Application to Program Follow Through in U.S. Education," unpublished Ph.D. thesis, Carnegie-Mellon University, School of Urban and Public Affairs, Pittsburgh, PA.

7. A. Charnes, W.W. Cooper and E. Rhodes (1979), "Short Communication: Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 3, p.339.

8. A. Charnes, W.W. Cooper and E. Rhodes (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Management Science* 27, pp.668-697.

9. A. Charnes and W.W. Cooper (1985), "Preface to Topics in Data Envelopment Analysis," *Annals of Operations Research* 2, pp.59-94

10. S. Bulla, W.W. Cooper, K.S. Park and D. Wilson (2000), *Journal of Propulsion and Power* 16, pp.431-439.

11. See Research Report CCS 532(Austin, Texas: University of Texas at Austin, Graduate School of Business, 1986) "Data Envelopment Analysis Approaches to Policy Evaluation and Management of Army Recruiting Activities I: Tradeoff between Joint Service and Army Advertising" by A. Charnes, W.W. Cooper, B. Golany, Major R. Halek, G. Klopp, E. Schmitz and Captain D. Thomas.

3 THE CCR MODEL AND PRODUCTION CORRESPONDENCE

3.1 INTRODUCTION

In this chapter, we relax the *positive* data set assumption. We shall instead assume the data are semipositive. That is, we assume that some (but not all) inputs and outputs are positive. This will allow us to deal with applications which involve zero data in inputs and/or outputs. The *production possibility set* composed of these input and output data (X, Y) will also be introduced. The *dual* problem of the CCR model will then be constructed and it will be shown that the dual problem evaluates efficiency based on a linear programming problem applied to the data set (X, Y) . The CCR-efficiency will be redefined, taking into account all input excesses and output shortfalls.

The observations that form the production possibility set are very fundamental in that they make it possible to assess the CCR model from a broader point of view and to extend this model to other models that will be introduced in succeeding chapters.

One version of a CCR model aims to minimize inputs while satisfying at least the given output levels. This is called the *input-oriented* model. There is another type of model called the *output-oriented* model that attempts to maximize outputs without requiring more of any of the observed input values. In the last section, the latter will be introduced along with a combination of the two models.

Computational aspects of the CCR model are also covered in this chapter. The computer code DEA-Solver accompanying this book (as mentioned in Problem 1.1 of Chapter 1) will be utilized on some of the problems we provide.

3.2 PRODUCTION POSSIBILITY SET

We have been dealing with the pairs of positive input and output vectors $(\mathbf{x}_j, \mathbf{y}_j)$ ($j = 1, \dots, n$) of n DMUs. In this chapter, the positive data assumption is relaxed. All data are assumed to be nonnegative but at least one component of every input and output vector is positive. We refer to this as *semipositive* with a mathematical characterization given by $\mathbf{x}_j \geq \mathbf{0}, \mathbf{x}_j \neq \mathbf{0}$ and $\mathbf{y}_j \geq \mathbf{0}, \mathbf{y}_j \neq \mathbf{0}$ for $j = 1, \dots, n$. Therefore, each DMU is supposed to have at least one positive value in both input and output. We will call a pair of such semipositive input $\mathbf{x} \in R^m$ and output $\mathbf{y} \in R^s$ an *activity* and express them by the notation (\mathbf{x}, \mathbf{y}) . The components of each such vector pair can be regarded as a semipositive orthant point in $(m + s)$ dimensional linear vector space in which the superscript m and s specify the number of dimensions required to express inputs and outputs, respectively. The set of feasible activities is called the *production possibility set* and is denoted by P . We postulate the following

Properties of P (the Production Possibility Set)

- (A1) The observed activities $(\mathbf{x}_j, \mathbf{y}_j)$ ($j = 1, \dots, n$) belong to P .
- (A2) If an activity (\mathbf{x}, \mathbf{y}) belongs to P , then the activity $(t\mathbf{x}, t\mathbf{y})$ belongs to P for any positive scalar t . We call this property the *constant returns-to-scale* assumption.
- (A3) For an activity (\mathbf{x}, \mathbf{y}) in P , any semipositive activity $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ with $\bar{\mathbf{x}} \geq \mathbf{x}$ and $\bar{\mathbf{y}} \leq \mathbf{y}$ is included in P . That is, any activity with input no less than \mathbf{x} in any component and with output no greater than \mathbf{y} in any component is feasible.
- (A4) Any semipositive linear combination of activities in P belongs to P .¹

Arranging the data sets in matrices $X = (\mathbf{x}_j)$ and $Y = (\mathbf{y}_j)$, we can define the production possibility set P satisfying (A1) through (A4) by

$$P = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}, \quad (3.1)$$

where $\boldsymbol{\lambda}$ is a semipositive vector in R^n .

Figure 3.1 shows a typical production possibility set in two dimensions for the single input and single output case, so that $m = 1$ and $s = 1$, respectively. In this example the possibility set is determined by B and the ray from the origin through B is the efficient frontier.

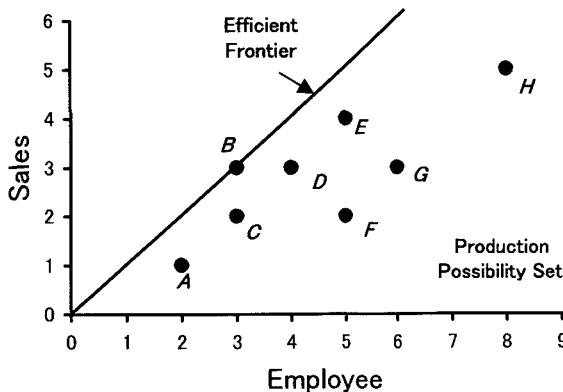


Figure 3.1. Production Possibility Set

3.3 THE CCR MODEL AND DUAL PROBLEM

Based on the matrix (X, Y) , the CCR model was formulated in the preceding chapter as an LP problem with row vector v for input multipliers and row vector u as output multipliers. These multipliers are treated as variables in the following LP problem ([Multiplier form]):

$$(LP_o) \quad \max_{v, u} \quad u y_o \quad (3.2)$$

$$\text{subject to} \quad v x_o = 1 \quad (3.3)$$

$$-v X + u Y \leq 0 \quad (3.4)$$

$$v \geq 0, \quad u \geq 0. \quad (3.5)$$

This is the same as (2.7)-(2.11), (LP_o) in the preceding chapter, which is now expressed in vector-matrix notation.

The dual problem² of (LP_o) is expressed with a real variable θ and a non-negative vector $\lambda = (\lambda_1, \dots, \lambda_n)^T$ of variables as follows ([Envelopment form]):

$$(DLP_o) \quad \min_{\theta, \lambda} \quad \theta \quad (3.6)$$

$$\text{subject to} \quad \theta x_o - X \lambda \geq 0 \quad (3.7)$$

$$Y \lambda \geq y_o \quad (3.8)$$

$$\lambda \geq 0. \quad (3.9)$$

Correspondences between the primal (LP_o) and the dual (DLP_o) constraints and variables are displayed in Table 3.1.

(DLP_o) has a feasible solution $\theta = 1, \lambda_o = 1, \lambda_j = 0 (j \neq o)$. Hence the optimal θ , denoted by θ^* , is not greater than 1. On the other hand, due to the nonzero (i.e., semipositive) assumption for the data, the constraint (3.8)

Table 3.1. Primal and Dual Correspondences

Constraint (LP_o)	Dual variable (DLP_o)		Constraint (DLP_o)	Primal variable (LP_o)
$v\mathbf{x}_o = 1$	θ		$\theta\mathbf{x}_o - X\boldsymbol{\lambda} \geq \mathbf{0}$	$v \geq \mathbf{0}$
$-vX + uY \leq \mathbf{0}$	$\boldsymbol{\lambda} \geq \mathbf{0}$		$Y\boldsymbol{\lambda} \geq \mathbf{y}_o$	$u \geq \mathbf{0}$

forces $\boldsymbol{\lambda}$ to be nonzero because $\mathbf{y}_o \geq \mathbf{0}$ and $\mathbf{y}_o \neq \mathbf{0}$. Hence, from (3.7), θ must be greater than zero. Putting this all together, we have $0 < \theta^* \leq 1$. Now we observe the relation between the production possibility set P and (DLP_o) . The constraints of (DLP_o) require the activity $(\theta\mathbf{x}_o, \mathbf{y}_o)$ to belong to P , while the objective seeks the minimum θ that reduces the input vector \mathbf{x}_o radially to $\theta\mathbf{x}_o$ while remaining in P . In (DLP_o) , we are looking for an activity in P that guarantees at least the output level \mathbf{y}_o of DMU_o in all components while reducing the input vector \mathbf{x}_o proportionally (radially) to a value as small as possible. Under the assumptions of the preceding section, it can be said that $(X\boldsymbol{\lambda}, Y\boldsymbol{\lambda})$ outperforms $(\theta\mathbf{x}_o, \mathbf{y}_o)$ when $\theta^* < 1$. With regard to this property, we define the input *excesses* $s^- \in R^m$ and the output *shortfalls* $s^+ \in R^s$ and identify them as “slack” vectors by:

$$s^- = \theta\mathbf{x}_o - X\boldsymbol{\lambda}, \quad s^+ = Y\boldsymbol{\lambda} - \mathbf{y}_o, \quad (3.10)$$

with $s^- \geq \mathbf{0}$, $s^+ \geq \mathbf{0}$ for any feasible solution $(\theta, \boldsymbol{\lambda})$ of (DLP_o) .

To discover the possible input excesses and output shortfalls, we solve the following two-phase LP problem.

Phase I

We solve (DLP_o) . Let the optimal objective value be θ^* . By the duality theorem of linear programming,³ θ^* is equal to the optimal objective value of (LP_o) and is the CCR-efficiency value, also called “Farrell Efficiency,” after M.J. Farrell (1957). See below. This value of θ^* is incorporated in the following Phase II extension of (DLP_o) .

Phase II

Using our knowledge of θ^* , we solve the following LP using $(\boldsymbol{\lambda}, s^-, s^+)$ as variables:

$$\max_{\boldsymbol{\lambda}, s^-, s^+} \omega = es^- + es^+ \quad (3.11)$$

$$\text{subject to } s^- = \theta^*\mathbf{x}_o - X\boldsymbol{\lambda} \quad (3.12)$$

$$s^+ = Y\boldsymbol{\lambda} - \mathbf{y}_o \quad (3.13)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \quad s^- \geq \mathbf{0}, \quad s^+ \geq \mathbf{0},$$

where $\mathbf{e} = (1, \dots, 1)$ (a vector of ones) so that $\mathbf{e}\mathbf{s}^- = \sum_{i=1}^m s_i^-$ and $\mathbf{e}\mathbf{s}^+ = \sum_{r=1}^s s_r^+$.

The objective of Phase II is to find a solution that maximizes the sum of input excesses and output shortfalls while keeping $\theta = \theta^*$.

We should note that we could replace the objective term in (3.11) with any weighted sum of input excesses and output shortfalls such as:

$$\omega = \mathbf{w}_x \mathbf{s}^- + \mathbf{w}_y \mathbf{s}^+, \quad (3.14)$$

where the weights \mathbf{w}_x and \mathbf{w}_y are positive row vectors. The modified objective function may result in a different optimal solution for Phase II. However, we can have the optimal $\omega^* > 0$ in (3.11) if and only if a nonzero value is also obtained when the objective in (3.11) is replaced with (3.14). Thus the objective in (3.11) will identify some nonzero slacks with inefficiency if and only if some nonzero (possibly different) slacks are identified with inefficiency in (3.14).

Definition 3.1 (Max-slack Solution, Zero-slack Activity)

An optimal solution $(\lambda^, \mathbf{s}^-, \mathbf{s}^+)$ of Phase II is called the max-slack solution. If the max-slack solution satisfies $\mathbf{s}^- = \mathbf{0}$ and $\mathbf{s}^+ = \mathbf{0}$, then it is called zero-slack.*

Definition 3.2 (CCR-Efficiency, Radial Efficiency, Technical Efficiency)

If an optimal solution $(\theta^, \lambda^*, \mathbf{s}^-, \mathbf{s}^+)$ of the two LPs above satisfies $\theta^* = 1$ and is zero-slack ($\mathbf{s}^- = \mathbf{0}$, $\mathbf{s}^+ = \mathbf{0}$), then the DMU_o is called CCR-efficient. Otherwise, the DMU_o is called CCR-inefficient, because*

- (i) $\theta^* = 1$
- (ii) All slacks are zero

must both be satisfied if full efficiency is to be attained.

The first of these two conditions is referred to as “radial efficiency.” It is also referred to as “technical efficiency” because a value of $\theta^* < 1$ means that all inputs can be simultaneously reduced without altering the mix (=proportions) in which they are utilized. Because $(1 - \theta^*)$ is the maximal proportionate reduction allowed by the production possibility set, any further reductions associated with nonzero slacks will necessarily change the input proportions. Hence the inefficiencies associated with any nonzero slack identified in the above two-phase procedure are referred to as “mix inefficiencies.” Other names are also used to characterize these two sources of inefficiency. For instance, the term “weak efficiency” is sometime used when attention is restricted to (i) in Definition 3.2. The conditions (i) and (ii) taken together describe what is also called “Pareto-Koopmans” or “strong” efficiency, which can be verbalized as follows,

Definition 3.3 (Pareto-Koopmans Efficiency)

A DMU is fully efficient if and only if it is not possible to improve any input or output without worsening some other input or output.

This last definition recognizes the contributions of the economists Vilfredo Pareto and Tjalling Koopmans. Implementable form was subsequently given to it by M.J. Farrell, another economist, who showed how to apply these concepts to observed data. However, Farrell was only able to carry his developments to a point which satisfied condition (i) but not condition (ii) in Definition 3.2. Hence he did not fully satisfy the conditions for Pareto-Koopmans efficiency but stopped short, instead, with what we just referred to as “weak efficiency” (also called “Farrell efficiency”) because nonzero slack, when present in any input or output, can be used to effect additional improvements without worsening any other input or output. Farrell, we might note, was aware of this shortcoming in his approach which he tried to treat by introducing new (unobserved) “points at infinity” but was unable to give his concept implementable form.⁴ In any case, this was all accomplished by Charnes, Cooper and Rhodes in a mathematical formulation that led to the two-phase procedure described above. Hence we also refer to this as “CCR-efficiency” when this procedure is utilized on empirical data to fulfill both (i) and (ii) in Definition 3.2.

We have already given a definition of CCR-efficiency in Chapter 2. For the data set (X, Y) under the assumption of semipositivity, we can also define CCR-efficiency by Definition 2.1 in that chapter. We now prove that the CCR-efficiency above, gives the same efficiency characterization as is obtained from Definition 2.1 in Chapter 2. This is formalized by:

Theorem 3.1 *The CCR-efficiency given in Definition 3.2 is equivalent to that given by Definition 2.1.*

Proof. First, notice that the vectors \mathbf{v} and \mathbf{u} of (LP_o) are dual multipliers corresponding to the constraints (3.7) and (3.8) of (DLP_o) , respectively. See Table 3.1. Now the following “complementary conditions”⁵ hold between any optimal solutions $(\mathbf{v}^*, \mathbf{u}^*)$ of (LP_o) and $(\lambda^*, s^{-*}, s^{+*})$ of (DLP_o) .

$$\mathbf{v}^* s^{-*} = 0 \text{ and } \mathbf{u}^* s^{+*} = 0. \quad (3.15)$$

This means that if any component of \mathbf{v}^* or \mathbf{u}^* is positive then the corresponding component of s^{-*} or s^{+*} must be zero, and conversely, with the possibility also allowed in which both components may be zero simultaneously.

Now we demonstrate that Definition 3.2 implies Definition 2.1

(i) If $\theta^* < 1$, then DMU_o is CCR-inefficient by Definition 2.1, since (LP_o) and (DLP_o) have the same optimal objective value θ^* .

(ii) If $\theta^* = 1$ and is not zero-slack ($s^{-*} \neq \mathbf{0}$ and/or $s^{+*} \neq \mathbf{0}$), then, by the complementary conditions above, the elements of \mathbf{v}^* or \mathbf{u}^* corresponding to the positive slacks must be zero. Thus, DMU_o is CCR-inefficient by Definition 2.1.

(iii) Lastly if $\theta^* = 1$ and zero-slack, then, by the “strong theorem of complementarity,”⁶ (LP_o) is assured of a positive optimal solution $(\mathbf{v}^*, \mathbf{u}^*)$ and hence DMU_o is CCR-efficient by Definition 2.1.

The reverse is also true by the complementary relation and the strong complementarity theorem between $(\mathbf{v}^*, \mathbf{u}^*)$ and (s^{-*}, s^{+*}) . \square

3.4 THE REFERENCE SET AND IMPROVEMENT IN EFFICIENCY

Definition 3.4 (Reference Set)

For an inefficient DMU_o , we define its reference set E_o , based on the max-slack solution as obtained in phases one and two, by

$$E_o = \{j \mid \lambda_j^* > 0\} \quad (j \in \{1, \dots, n\}). \quad (3.16)$$

An optimal solution can be expressed as

$$\begin{aligned} \theta^* x_o &= \sum_{j \in E_o} x_j \lambda_j^* + s^{-*} \\ y_o &= \sum_{j \in E_o} y_j \lambda_j^* - s^{+*}. \end{aligned} \quad (3.17)$$

This can be interpreted as follows,

$$x_o \geq \theta^* x_o - s^{-*} = \sum_{j \in E_o} x_j \lambda_j^*$$

which means

$$\begin{aligned} x_o &\geq \text{technical} - \text{mix inefficiency} \\ &= \text{a positive combination of observed input values.} \end{aligned} \quad (3.18)$$

Also

$$y_o \leq y_o + s^{+*} = \sum_{j \in E_o} y_j \lambda_j^*$$

means

$$\begin{aligned} y_o &\leq \text{observed outputs} + \text{shortfalls} \\ &= \text{a positive combination of observed output values.} \end{aligned} \quad (3.19)$$

These relations suggest that the efficiency of (x_o, y_o) for DMU_o can be improved if the input values are reduced radially by the ratio θ^* and the input excesses recorded in s^{-*} are eliminated. Similarly efficiency can be attained if the output values are augmented by the output shortfalls in s^{+*} . Thus, we have a method for improving an inefficient DMU that accords with Definition 3.2. The gross input improvement Δx_o and output improvement Δy_o can be calculated from:

$$\Delta x_o = x_o - (\theta^* x_o - s^{-*}) = (1 - \theta^*) x_o + s^{-*} \quad (3.20)$$

$$\Delta y_o = s^{+*}. \quad (3.21)$$

Hence, we have a formula for improvement, which is called *the CCR projection*:⁷

$$\hat{x}_o = x_o - \Delta x_o = \theta^* x_o - s^{-*} \leq x_o \quad (3.22)$$

$$\hat{y}_o = y_o + \Delta y_o = y_o + s^{+*} \geq y_o. \quad (3.23)$$

However, note that there are other formulae for improvement as will be described later.

In Theorems 3.2, 3.3 and 3.4 in the next section, we will show that the improved activity (\hat{x}_o, \hat{y}_o) projects DMU_o into the reference set E_o and any nonnegative combination of DMUs in E_o is efficient.

3.5 THEOREMS ON CCR-EFFICIENCY

Theorem 3.2 *The improved activity (\hat{x}_o, \hat{y}_o) defined by (3.22) and (3.23) is CCR-efficient.*

Proof. The efficiency of (\hat{x}_o, \hat{y}_o) is evaluated by solving the LP problem below:

$$(DLP_e) \quad \min_{\theta, \lambda, s^-, s^+} \theta \quad (3.24)$$

$$\text{subject to} \quad \theta \hat{x}_o - X\lambda - s^- = \mathbf{0} \quad (3.25)$$

$$Y\lambda - s^+ = \hat{y}_o \quad (3.26)$$

$$\lambda \geq \mathbf{0}, \quad s^- \geq \mathbf{0}, \quad s^+ \geq \mathbf{0}. \quad (3.27)$$

Let an optimal (max-slack) solution for (DLP_e) be $(\hat{\theta}, \hat{\lambda}, \hat{s}^-, \hat{s}^+)$. By inserting the formulae (3.22) and (3.23) into the constraints, we have

$$\begin{aligned} \hat{\theta}\theta^*x_o &= X\hat{\lambda} + \hat{s}^- + \hat{\theta}s^{-*} \\ y_o &= Y\hat{\lambda} - \hat{s}^+ - s^{+*}. \end{aligned}$$

Now we can also write this solution as

$$\begin{aligned} \tilde{\theta}x_o &= X\hat{\lambda} + \tilde{s}^- \\ y_o &= Y\hat{\lambda} - \tilde{s}^+ \end{aligned}$$

where $\tilde{\theta} = \hat{\theta}\theta^*$ and $\tilde{s}^- = \hat{s}^- + \hat{\theta}s^{-*} \geq \mathbf{0}$, $\tilde{s}^+ = \hat{s}^+ + s^{+*} \geq \mathbf{0}$. However, θ^* is part of an optimal solution so we must have $\tilde{\theta} = \hat{\theta}\theta^* = \theta^*$ so $\hat{\theta} = 1$. Furthermore, with $\hat{\theta} = 1$ we have

$$e\tilde{s}^- + e\tilde{s}^+ = (e\hat{s}^- + es^{-*}) + (e\hat{s}^+ + es^{+*}) \leq es^{-*} + es^{+*}$$

since $es^{-*} + es^{+*}$ is maximal. It follows that we must have $e\hat{s}^- + e\hat{s}^+ = \mathbf{0}$ which implies that all components of \hat{s}^- and \hat{s}^+ are zero. Hence conditions (i) and (ii) of Definition 3.2 are both satisfied and CCR-efficiency is achieved as claimed. \square

Corollary 3.1 (Corollary to Theorem 3.2)

The point with coordinates \hat{x}_o, \hat{y}_o defined by (3.22) and (3.23), viz.,

$$\hat{x}_o = \theta^*x_o - s^{-*} = \sum_{j \in E_o} x_j \lambda_j^* \quad (3.28)$$

$$\hat{y}_o = y_o + s^{+*} = \sum_{j \in E_o} y_j \lambda_j^* \quad (3.29)$$

is the point on the efficient frontier used to evaluate the performance of DMU_o .

In short the CCR projections identify the point either as a positive combination of other DMUs with $\mathbf{x}_o \geq \hat{\mathbf{x}}_o$ and $\hat{\mathbf{y}}_o \geq \mathbf{y}_o$ unless $\theta^* = 1$ and all slacks are zero in which case $\mathbf{x}_o = \hat{\mathbf{x}}_o$ and $\hat{\mathbf{y}}_o = \mathbf{y}_o$ so the operation in (3.28) performed on the observation for DMU_o identifies a new DMU positioned on the efficient frontier. Conversely, the point associated with the thus generated DMU evaluates the performance of DMU_o as exhibiting input excesses $\mathbf{x}_o - \hat{\mathbf{x}}_o$ and output shortfalls $\hat{\mathbf{y}}_o - \mathbf{y}_o$.

We notice that the improvement by the formulae (3.22) and (3.23) should be achieved by using the *max-slack* solution. If we do it based on another (not max-slack) solution, the improved activity $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ is not necessarily CCR-efficient. This is shown by the following example with 4 DMUs A, B, C and D , each with 3 inputs x_1, x_2 and x_3 and all producing 1 output in amount $y = 1$.

	A	B	C	D
x_1	2	2	2	1
x_2	1	1.5	2	1
x_3	1	1	1	2
y	1	1	1	1

Here the situation for DMU C is obvious. We can observe, for instance, that DMU C has two possible slacks in x_2 , 1 against A and 0.5 against B . If we choose B as the reference set of C , then the improved activity coincides with B , which still has a slack of 0.5 in x_2 against A . Therefore, the improved activity is not CCR-efficient. However, if we improve C by using the max-slack solution, then we move to the CCR-efficient DMU A directly.

Lemma 3.1 *For the improved activity $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$, there exists an optimal solution $(\hat{\mathbf{v}}_o, \hat{\mathbf{u}}_o)$ for the problem (LP_e) , which is dual to (DLP_e) , such that*

$$\begin{aligned} \hat{\mathbf{v}}_o &> \mathbf{0} \text{ and } \hat{\mathbf{u}}_o > \mathbf{0} \\ \hat{\mathbf{v}}_o \mathbf{x}_j &= \hat{\mathbf{u}}_o \mathbf{y}_j \quad (j \in E_o) \end{aligned} \quad (3.30)$$

$$\hat{\mathbf{v}}_o X \geq \hat{\mathbf{u}}_o Y \quad (3.31)$$

Proof. Since $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ is zero-slack, the strong theorem of complementarity means that there exists a positive optimal solution $(\hat{\mathbf{v}}_o, \hat{\mathbf{u}}_o)$ for (LP_e) . The equality (3.30) is the complementarity condition between primal-dual optimal solutions. The inequality (3.31) is a part of the constraints of (LP_e) . \square

Theorem 3.3 *The DMUs in E_o as defined in (3.16) are CCR-efficient.*

Proof. As described in Lemma 3.1, there exists a positive multiplier $(\hat{\mathbf{v}}_o, \hat{\mathbf{u}}_o)$. These vectors also satisfy

$$\hat{\mathbf{v}}_o \mathbf{x}_j = \hat{\mathbf{u}}_o \mathbf{y}_j \quad (j \in E_o) \quad (3.32)$$

$$\hat{\mathbf{v}}_o X \geq \hat{\mathbf{u}}_o Y \quad (3.33)$$

For each $j \in E_o$, we can adjust $(\hat{\mathbf{v}}_o, \hat{\mathbf{u}}_o)$ by using a scalar multiplier so that the relation $\hat{\mathbf{v}}_o \mathbf{x}_j = \hat{\mathbf{u}}_o \mathbf{y}_j = 1$ holds, while keeping (3.33) satisfied. Thus, activity $(\mathbf{x}_j, \mathbf{y}_j)$ is CCR-efficient by Definition 2.1. (A proof via the envelopment model may be found in Problem 7.5 of Chapter 7, and this proof comprehends the BCC as well as the CCR model). \square

Theorem 3.4 *Any semipositive combination of DMUs in E_o is CCR-efficient.*

Proof. Let the combined activity be

$$\mathbf{x}_c = \sum_{j \in E_o} c_j \mathbf{x}_j \text{ and } \mathbf{y}_c = \sum_{j \in E_o} c_j \mathbf{y}_j \text{ with } c_j \geq 0 \quad (j \in E_o). \quad (3.34)$$

The multiplier $(\hat{\mathbf{v}}_o, \hat{\mathbf{u}}_o)$ in Lemma 3.1 satisfies

$$\hat{\mathbf{v}}_o \mathbf{x}_c = \hat{\mathbf{u}}_o \mathbf{y}_c \quad (3.35)$$

$$\hat{\mathbf{v}}_o \mathbf{X} \geq \hat{\mathbf{u}}_o \mathbf{Y} \quad (3.36)$$

$$\hat{\mathbf{v}}_o > \mathbf{0}, \hat{\mathbf{u}}_o > \mathbf{0} \quad (3.37)$$

Thus, $(\mathbf{x}_c, \mathbf{y}_c)$ is CCR-efficient by Definition 2.1. \square

3.6 COMPUTATIONAL ASPECTS OF THE CCR MODEL

In this section, we discuss computational procedures for solving linear programs for the CCR model in detail. Readers who are not familiar with the terminology of linear programs and are not interested in the computational aspects of DEA can skip this section.

3.6.1 Computational Procedure for the CCR Model

As described in Section 3.3, the computational scheme of the CCR model for DMU_o results in the following two stage LP problem.

$$(DLP_o) \quad \min \theta \quad (3.38)$$

$$\text{Phase I objective} \quad \min -\mathbf{es}^- - \mathbf{es}^+ \quad (3.39)$$

$$\text{subject to} \quad \theta \mathbf{x}_o = \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- \quad (3.40)$$

$$\mathbf{y}_o = \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ \quad (3.41)$$

$$\theta \geq 0, \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, \quad (3.42)$$

where Phase II replaces the variable θ with a fixed value of $\min \theta = \theta^*$.

Using the usual LP notation, we now rewrite (DLP_o) as follows,

$$(DLP'_o) \quad \min z_1 = \mathbf{c}\mathbf{x} \quad (3.43)$$

$$\text{Phase II objective} \quad \min z_2 = \mathbf{d}\mathbf{x} \quad (3.44)$$

$$\text{subject to} \quad \mathbf{Ax} = \mathbf{b} \quad (3.45)$$

$$\mathbf{x} \geq \mathbf{0}, \quad (3.46)$$

where \mathbf{c} and \mathbf{d} are row vectors.

The correspondences between (DLP_o) and (DLP'_o) are:

$$\mathbf{x} = (\theta, \lambda^T, s^{-T}, s^{+T})^T \quad (3.47)$$

$$\mathbf{c} = (1, \mathbf{0}, \mathbf{0}, \mathbf{0}) \quad (3.48)$$

$$\mathbf{d} = (0, \mathbf{0}, -\mathbf{e}, -\mathbf{e}) \quad (3.49)$$

$$A = \begin{pmatrix} \mathbf{x}_o & -X & -I & O \\ \mathbf{0} & Y & O & -I \end{pmatrix} \quad (3.50)$$

$$\mathbf{b} = \begin{pmatrix} \mathbf{0} \\ \mathbf{y}_o \end{pmatrix}, \quad (3.51)$$

where \mathbf{e} is the vector with all elements unity. See the discussion immediately following (3.13).

Phase I

First, we solve the LP problem with the Phase I objective. Letting an optimal basis be B we use this matrix B as follows. First we compute several values, as noted below, where R is the nonbasic part of the matrix A and the superscript B (or R) shows the columns of A corresponding to B (or R).

$$\text{basic solution} \quad \mathbf{x}^B = \bar{\mathbf{b}} = B^{-1}\mathbf{b} \quad (3.52)$$

$$\text{simplex multiplier} \quad \boldsymbol{\pi} = \mathbf{c}^B B^{-1} \quad (3.53)$$

$$\text{Phase I simplex criterion} \quad \mathbf{p}^R = \boldsymbol{\pi}R - \mathbf{c}^R \quad (3.54)$$

$$\text{Phase II simplex criterion} \quad \mathbf{q}^R = \boldsymbol{\pi}R - \mathbf{d}^R, \quad (3.55)$$

where B^{-1} is the inverse of B , $\boldsymbol{\pi}$ is a vector of “multipliers” derived (from the data) as in (3.53), and \mathbf{p}^R and \mathbf{q}^R are referred to as “reduced costs.”

From optimality of the basis B , we have:

$$\bar{\mathbf{b}} \geq \mathbf{0}, \quad (3.56)$$

because the conditions for non-negativity are satisfied by \mathbf{x}^B in (3.52) and

$$\mathbf{p}^R \leq \mathbf{0}, \quad (3.57)$$

as required for optimality in (3.54) and, for the columns of the basis B , we have: $\mathbf{p}^B = \boldsymbol{\pi}B - \mathbf{c}^B = \mathbf{0}$.

Phase II

We exclude the columns with $p_j < 0$ in the optimal simplex tableau at the end of Phase I from the tableaux for further consideration. The remainder is called the *restricted problem* or *restricted tableau* and dealt with in the next computational step. At Phase II, we solve the LP problem with the second objective. If the simplex criterion of Phase II objective satisfies $\mathbf{q}^R \leq \mathbf{0}$, we then halt the iterations. The basic solution thus obtained is a max-slack solution. This is the procedure for finding θ^* and a max-slack solution $(\mathbf{s}^{-*}, \mathbf{s}^{+*})$ for DMU_o . As has been already pointed out, (LP_o) in (3.2)-(3.5) is dual to (DLP_o) in (3.6)-(3.9), and we will discuss its solution procedure in the next section.

3.6.2 Data Envelopment Analysis and the Data

The characterizations given by (3.52)-(3.55) represent borrowings from the terminology of linear programming in which the components of π are referred to as “simplex multipliers” because they are associated with a use of the simplex method which generates such multipliers in the course of solving LP problems such as (DLP'_o) . “Simplex” and “dual simplex” methods are also used in DEA. We therefore refer to the problem (LP_o) as being in “multiplier form.” See (3.2)-(3.5). Problem (DLP_o) is then said to be in “envelopment form.” This is the source of the name “Data Envelopment Analysis.”

Reference to Figure 3.2 in the next section will help to justify this usage by noting, first, that all data are inside the frontier stretching from F through R . Also, at least one observation is touching the frontier. Hence, in the manner of an envelope all data are said to be “enveloped” by the frontier which achieves its importance because it is used to analyze and evaluate the performances of DMUs associated with observations like those portrayed in Figure 3.2. More compactly, we say that DEA is used to evaluate the performances of each observation relative to the frontier that envelops all of the observations.

3.6.3 Determination of Weights (=Multipliers)

The simplex multipliers in π at the end of Phase I are associated with an optimal solution of its dual problem (LP'_o) ⁸ as given in (3.2)-(3.5). In fact, π is an $(m + s)$ vector in which the first m components assign optimal weights v^* to the inputs and the remaining s components assign optimal weights u^* to the outputs. By observing the dual problem of (DLP'_o) in (3.38)-(3.42), it can be shown that v^* and u^* satisfy (3.3) through (3.5). The symbol I , for the identity matrix in A of (3.50), relates to the (input and output) slack. Let the simplex criteria to be applied to these columns be represented by “pricing vectors” p^{s^-} and p^{s^+} . These vectors relate to v^* and u^* in the following way:

$$v^* = -p^{s^-} \quad (\geq 0) \quad (3.58)$$

$$u^* = -p^{s^+} \quad (\geq 0). \quad (3.59)$$

3.6.4 Reasons for Solving the CCR Model Using the Dual (DLP_o)

It is not advisable to solve (LP_o) directly. The reasons are:

- (1) The computational effort of LP is apt to grow in proportion to powers of the number of constraints. Usually in DEA, n , the number of DMUs is considerably larger than $(m + s)$, the number of inputs and outputs and hence it takes more time to solve (LP_o) which has n constraints than to solve (DLP_o) which has $(m + s)$ constraints. In addition, since the memory size needed for keeping the basis (or its inverse) is the square of the number of constraints, (DLP_o) is better fitted for memory saving purposes.
- (2) We cannot find the pertinent max-slack solution by solving (LP_o) .
- (3) The interpretations of (DLP_o) are more straightforward because the solutions are characterized as inputs and outputs that correspond to the original

data whereas the multipliers provided by solutions to (LP_o) represent evaluations of these observed values. These values are also important, of course, but they are generally best reserved for supplementary analyses after a solution to (DLP_o) is achieved.

3.7 EXAMPLE

We will apply (DLP_o) to a sample problem and comment on the results. For this purpose we use Example 3.1 as shown in Table 3.2, which is Example 2.2 with an added activity G . All computations were done using the DEA-Solver which comes with this book and the results are stored in the Excel 97 Workbook “Sample-CCR-I.xls” in the sample file. So, readers interested in using DEA-Solver should try to trace the results by opening this file or running DEA-Solver for this problem.

Table 3.2. Example 3.1

	DMU	A	B	C	D	E	F	G
Input	x_1	4	7	8	4	2	10	3
	x_2	3	3	1	2	4	1	7
Output	y	1	1	1	1	1	1	1

(1) (DLP_A) for A is:

$$\text{Phase I } \min \theta$$

$$\text{Phase II } \min -s_1^- - s_2^- - s^+$$

subject to

$$4\theta - 4\lambda_A - 7\lambda_B - 8\lambda_C - 4\lambda_D - 2\lambda_E - 10\lambda_F - 3\lambda_G - s_1^- = 0$$

$$3\theta - 3\lambda_A - 3\lambda_B - \lambda_C - 2\lambda_D - 4\lambda_E - \lambda_F - 7\lambda_G - s_2^- = 0$$

$$\lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F + \lambda_G - s^+ = 1$$

where the variables are restricted to nonnegative values in the vectors λ , s^- and s^+ .

The optimal solution for (DLP_A) is:

$$\theta^* = 0.8571 \quad (\text{in Worksheet ‘‘Sample-CCR-I.Score’’})$$

$$\lambda_D^* = 0.7143, \lambda_E^* = 0.2857, \text{ other } \lambda_j^* = 0 \quad (\text{in Worksheet ‘‘Sample-CCR-I.Score’’})$$

$$s_1^{-*} = s_2^{-*} = s^{+*} = 0 \quad (\text{in Worksheet ‘‘Sample-CCR-I.Slack’’}).$$

Since $\lambda_D^* > 0$ and $\lambda_E^* > 0$, the reference set for A is

$$E_A = \{D, E\}.$$

And $\lambda_D^* = 0.7143$, $\lambda_E^* = 0.2857$ show the proportions contributed by D and E to the point used to evaluate A . Hence A is technically inefficient. No mix inefficiencies are present because all slacks are zero. Thus removal of all inefficiencies is achieved by reducing all inputs by 0.1429 or, approximately, 15% of their observed values. In fact, based on this reference set and λ^* , we can express the input and output values needed to bring A into efficient status as

$$\begin{aligned} 0.8571 \times (\text{Input of } A) &= 0.7143 \times (\text{Input of } D) + 0.2857 \times (\text{Input of } E) \\ (\text{Output of } A) &= 0.7143 \times (\text{Output of } D) + 0.2857 \times (\text{Output of } E). \end{aligned}$$

From the magnitude of coefficients on the right hand side, A has more similarity to D than E . A can be made efficient either by using these coefficients, $\lambda_D^* = 0.7143$, $\lambda_E^* = 0.2857$ or by reducing both of its inputs—viz., by reducing the input value radially in the ratio 0.8571. It is the latter (radial contraction) that is used in DEA-Solver. Thus, as seen in Worksheet “Sample-CCR-I.Projection,” the CCR-projection of (3.32) and (3.33) is achieved by,

$$\begin{aligned} \hat{x}_1 &\leftarrow \theta^* x_1 = 0.8571 \times 4 = 3.4286 \quad (14.29\% \text{ reduction}) \\ \hat{x}_2 &\leftarrow \theta^* x_2 = 0.8571 \times 3 = 2.5714 \quad (14.29\% \text{ reduction}) \\ \hat{y} &\leftarrow y = 1 \quad (\text{no change}). \end{aligned}$$

The optimal solution for the multiplier problem (LPA) can be found in Worksheet “Sample-CCR-I.Weight” as follows,

$$v_1^* = 0.1429, \quad v_2^* = 0.1429, \quad u^* = 0.8571.$$

This solution satisfies constraints (3.3)-(3.5) and maximizes the objective in (3.2), i.e., $u^*y = 0.8571 \times 1 = 0.8571 = \theta^*$ in the optimal objective value of (DLP_A). (See Problem 3.3 for managerial roles of these optimal multipliers (weights) for improving A .)

The Worksheet “Sample-CCR-I.WeightedData” includes optimal weighted inputs and output, i.e.,

$$\begin{aligned} v_1^* x_1 &= 0.1429 \times 4 = 0.5714 \\ v_2^* x_2 &= 0.1429 \times 3 = 0.4286 \\ u^* y &= 0.8571 \times 1 = 0.8571. \end{aligned}$$

The sum of the first two terms is 1 which corresponds to the constraint (3.3). The last term is the optimal objective value in this single output case.

(2) Moving from DMU A to DMU B , the optimal solution for B is:

$$\begin{aligned} \theta^* &= 0.6316 \\ \lambda_A^* &= \lambda_B^* = 0, \quad \lambda_C^* = 0.1053, \quad \lambda_D^* = 0.8947, \quad \lambda_E^* = \lambda_F^* = \lambda_G^* = 0 \\ s_1^- &= s_2^- = s^+ = 0 \\ v_1^* &= 0.0526, \quad v_2^* = 0.2105, \quad u^* = 0.6316. \end{aligned}$$

Since $\lambda_C^* > 0$, $\lambda_D^* > 0$, the reference set for B is:

$$E_B = \{C, D\}.$$

B can be expressed as:

$$\begin{aligned} 0.6316 \times (\text{Input of } B) &= 0.1053 \times (\text{Input of } C) + 0.8947 \times (\text{Input of } D) \\ (\text{Output of } B) &= 0.1053 \times (\text{Output of } C) + 0.8947 \times (\text{Output of } D). \end{aligned}$$

That is, it can be expressed in ratio form, as shown on the left or as a nonnegative combination of $\lambda_j^* > 0$ values as shown on the right. Using the expression on the left, the CCR-projection for B is

$$\begin{aligned} \hat{x}_1 &\leftarrow \theta^* x_1 = 0.6316 \times 7 = 4.4211 \quad (36.84\% \text{ reduction}) \\ \hat{x}_2 &\leftarrow \theta^* x_2 = 0.6316 \times 3 = 1.8974 \quad (36.84\% \text{ reduction}) \\ \hat{y} &\leftarrow y = 1 \quad (\text{no change}). \end{aligned}$$

(3) C , D and E

These 3 DMUs are found to be efficient. (See Worksheet “Sample-CCR-I.Score.”)

(4) The optimal solution of the LP problem for F is:

$$\begin{aligned} \theta^* &= 1 \\ \lambda_A^* &= \lambda_B^* = 0, \lambda_C^* = 1, \lambda_D^* = \lambda_E^* = \lambda_F^* = \lambda_G^* = 0 \\ s_1^{-*} &= 2, s_2^{-*} = s^{+*} = 0 \\ v_1^* &= 0, v_2^* = 1, u^* = 1 \end{aligned}$$

where, again, the last set of values refer to the multiplier problem. For the envelopment model we have $\lambda_C^* > 0$ as the only positive value of λ . Hence the reference set for F is:

$$E_F = \{C\}.$$

Considering the excess in Input 1 ($s_1^{-*} = 2$), F can be expressed as:

$$\begin{aligned} (\text{Input 1 of } F) &= (\text{Input 1 of } C) + 2 \\ (\text{Input 2 of } F) &= (\text{Input 2 of } C) \\ (\text{Output of } F) &= (\text{Output of } C). \end{aligned}$$

Although F is “radial-efficient,” it is nevertheless “CCR-inefficient” due to this excess (mix inefficiency) associated with $s_1^{-*} = 2$. Thus the performance of F can be improved by subtracting 2 units from Input 1. This can be accomplished by subtracting 2 units from input 1 on the left and setting $s_1^{-*} = 0$ on the right without worsening any other input and output. Hence condition (ii) in Definition 3.2 is not satisfied until this is done, so F did not achieve Pareto-Koopmans efficiency in its performance. See Definition 3.3.

(5) The optimal solution of the LP problem for G is:

$$\begin{aligned}\theta^* &= 0.6667 \\ \lambda_A^* = \lambda_B^* = \lambda_C^* = \lambda_D^* &= 0, \lambda_E^* = 1, \lambda_F^* = \lambda_G^* = 0 \\ s_1^{-*} &= 0, s_2^{-*} = 0.6667, s^{+*} = 0 \\ v_1^* &= 0.3333, v_2^* = 0, u^* = 0.6667\end{aligned}$$

Since $\lambda_E^* > 0$, the reference set for G is:

$$E_G = \{E\}.$$

Considering the excess in Input 2 ($s_2^{-*} = 0.6667$), G can be expressed by:

$$\begin{aligned}0.6667 \times (\text{Input 1 of } G) &= (\text{Input 1 of } E) \\ 0.6667 \times (\text{Input 2 of } G) &= (\text{Input 2 of } E) + 0.6667 \\ (\text{Output of } G) &= (\text{Output of } E).\end{aligned}$$

One plan for the improvement of G is to reduce all input values by multiplying them by 0.6667 and further subtracting 0.6667 from Input 2. When this is done the thus altered values coincide with the coordinates of E . Geometrically then, the CCR-projection for G is

$$\begin{aligned}\hat{x}_1 &\leftarrow \theta^* x_1 - s_1^{-*} = 0.6667 \times 3 - 0 = 2 \quad (33.33\% \text{ reduction}) \\ \hat{x}_2 &\leftarrow \theta^* x_2 - s_2^{-*} = 0.6667 \times 7 - 0.6667 = 4 \quad (42.86\% \text{ reduction}) \\ \hat{y} &\leftarrow y = 1 \quad (\text{no change}),\end{aligned}$$

where $\hat{x}_1 = 2$, $\hat{x}_2 = 4$, $\hat{y} = 1$ which values are the same as for E in Table 3.2.

The above observations are illustrated by Figure 3.2, which depicts Input 1 and Input 2 values of all DMUs. Since the output value is 1 for all the DMUs, we can compare their efficiencies via the input values.

The efficient frontier consists of the bold line CDE and the production possibility set is the region enclosed by this efficient frontier line plus the vertical line going up from E and the horizontal line extending to the right from C . Let the intersection point of OA and DE be Q . The activity Q has input proportional to that of A (4, 3) and is the least input value point on OA in the production possibility set.

$$\frac{OQ}{OA} = 0.8571$$

corresponds to radial (or ratio) efficiency of A .⁹ Also, we have:

$$\begin{aligned}\text{Input 1 of } Q &= 0.8571 \times 4 \quad (\text{Input 1 of } A) = 3.428 = \hat{x}_1 \\ \text{Input 2 of } Q &= 0.8571 \times 3 \quad (\text{Input 2 of } A) = 2.571 = \hat{x}_2.\end{aligned}$$

However, Q is the point that divides D and E in the ratio 0.7143 to 0.2857.

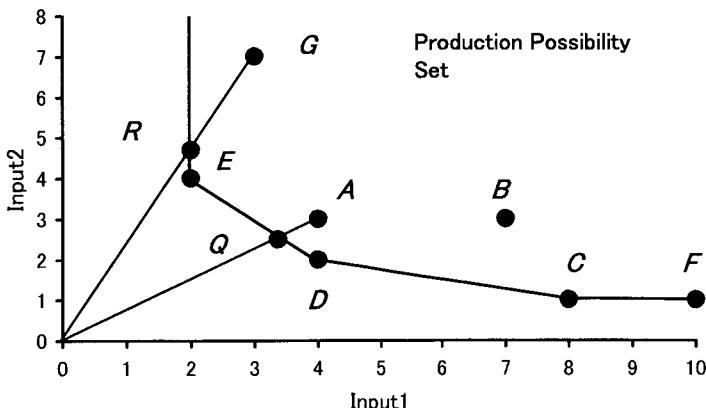


Figure 3.2. Example 3.1

Hence its values are calculated again as:

$$\begin{aligned} \text{Input 1 of } Q &= 0.7143 \times 4 \text{ (Input 1 of } D) + 0.2857 \times 2 \text{ (Input 1 of } E) = 3.428 \\ \text{Input 2 of } Q &= 0.7143 \times 2 \text{ (Input 2 of } D) + 0.2857 \times 4 \text{ (Input 2 of } E) = 2.571, \end{aligned}$$

where $\lambda_D^* = 0.7143$ and $\lambda_E^* = 0.2857$. Comparing these results we see that the coordinates of Q , the DMU used to evaluate A , can be derived in either of these two ways. There are also still more ways of effecting such improvements, as we will later see when changes in mix proportions are permitted.

This brings us to the role of nonzero slacks which we can illustrate with F and G . From Figure 3.2, it is evident that if we reduce Input 1 of F by the nonzero slack value of 2, then F coincides with C and is efficient. The presence of this nonzero slack means that F is not Pareto-Koopmans efficient even though its radial value is $\theta^* = 1$.

As a further illustration of the two conditions for efficiency in Definition 3.2 we turn to the evaluation of G . The CCR-efficiency of G is calculated by $OR/OG = 0.6667$. Thus G is not radially (or weakly) efficient as evaluated at R . However R is not efficient. We can make it efficient by further reducing Input 2 by 0.6667 and shifting to E as we have seen in the CCR-projection of G . Hence G fails both of the conditions specified in Definition 3.2.

Table 3.3 summarizes the results obtained by applying DEA-Solver to all of the data in Table 3.2. Only C , D and E are fully efficient. A and B fail because $\theta^* < 1$. Their intersection with the frontier gives their radial (weak) inefficiency score with zero slacks because they intersect an efficient portion of the frontier radially. F has a value of $\theta^* = 1$ because it is on the frontier. This portion of frontier is not efficient, however, as evidenced by the nonzero slack for F under s_1^- in Table 3.3. Finally, G fails to be efficient both because $\theta^* < 1$ and nonzero slacks are involved in its evaluation by E , a point on the efficient portion of the frontier.

Table 3.3. Results of Example 3.1

DMU	CCR-Eff θ^*	Ref Set		Excess	Shortfall
				s_1^-	s_2^-
A	0.8571	D	E	0	0
B	0.6316	C	D	0	0
C	1.0000	C		0	0
D	1.0000	D		0	0
E	1.0000	E		0	0
F	1.0000	C		2	0
G	0.6667	E		0	.6667

3.8 THE OUTPUT-ORIENTED MODEL

Up to this point, we have been dealing mainly with a model whose objective is to minimize inputs while producing at least the given output levels. This type of model is called *input-oriented*. There is another type of model that attempts to maximize outputs while using no more than the observed amount of any input. This is referred to as the *output-oriented* model, formulated as:

$$(DLPO_o) \quad \max_{\eta, \mu} \quad \eta \quad (3.60)$$

$$\text{subject to} \quad x_o - X\mu \geq 0 \quad (3.61)$$

$$\eta y_o - Y\mu \leq 0 \quad (3.62)$$

$$\mu \geq 0. \quad (3.63)$$

An optimal solution of $(DLPO_o)$ can be derived directly from an optimal solution of the input-oriented CCR model given in (3.6)-(3.9) as follows. Let us define

$$\lambda = \mu/\eta, \quad \theta = 1/\eta. \quad (3.64)$$

Then $(DLPO_o)$ becomes

$$(DLP_o) \quad \min_{\theta, \lambda} \quad \theta$$

$$\text{subject to} \quad \theta x_o - X\lambda \geq 0$$

$$y_o - Y\lambda \leq 0$$

$$\lambda \geq 0,$$

which is the input-oriented CCR model. Thus, an optimal solution of the output-oriented model relates to that of the input-oriented model via:

$$\eta^* = 1/\theta^*, \quad \mu^* = \lambda^*/\theta^*. \quad (3.65)$$

The slack (t^- , t^+) of the output-oriented model is defined by:

$$\begin{aligned} X\mu + t^- &= x_o \\ Y\mu - t^+ &= \eta y_o. \end{aligned}$$

These values are also related to the input-oriented model via

$$t^{-*} = s^{-*}/\theta^*, \quad t^{+*} = s^{+*}/\theta^*. \quad (3.66)$$

Now, $\theta^* \leq 1$, so returning to (3.64), η^* satisfies

$$\eta^* \geq 1. \quad (3.67)$$

The higher the value of η^* , the less efficient the DMU. θ^* expresses the input reduction rate, while η^* describes the output enlargement rate. From the above relations, we can conclude that an input-oriented CCR model will be efficient for any DMU if and only if it is also efficient when the output-oriented CCR model is used to evaluate its performance.

The dual problem of ($DLPO_o$) is expressed in the following model, with components of the vectors p and q serving as variables.

$$(LPO_o) \quad \min_{p,q} \quad p x_o \quad (3.68)$$

$$\text{subject to} \quad q y_o = 1 \quad (3.69)$$

$$-pX + qY \leq 0 \quad (3.70)$$

$$p \geq 0, \quad q \geq 0. \quad (3.71)$$

On the multiplier side we have:

Theorem 3.5 *Let an optimal solution of (LP_o) be (v^*, u^*) , then an optimal solution of the output-oriented model (LPO_o) is obtained from*

$$p^* = v^*/\theta^*, \quad q^* = u^*/\theta^*. \quad (3.72)$$

Proof. It is clear that (p^*, q^*) is feasible for (LPO_o). Its optimality comes from the equation below.

$$p^* x_o = v^* x_o / \theta^* = \eta^*. \quad (3.73)$$

□

Thus, the solution of the output-oriented CCR model may be obtained from that of the input oriented CCR model. The improvement using this model is expressed by:

$$\hat{x}_o \Leftarrow x_o - t^{-*} \quad (3.74)$$

$$\hat{y}_o \Leftarrow \eta^* y_o + t^{+*}. \quad (3.75)$$

Carrying this a stage further we note that (LPO_o) is equivalent to the following fractional programming problem:

$$\min_{\pi, \rho} \frac{\pi x_o}{\rho y_o} \quad (3.76)$$

$$\text{subject to } \frac{\pi x_j}{\rho y_j} \geq 1 \quad (j = 1, \dots, n) \quad (3.77)$$

$$\pi \geq 0, \quad \rho \geq 0. \quad (3.78)$$

That is, we exchanged the numerator and the denominator of (2.3) and (2.4) as given in Chapter 2 and minimized the objective function. It is therefore quite natural that the solutions found to be linked by a simple rule. This mathematical transformation does not imply that there is no managerial significance to be assigned to the choice of models since, *inter alia*, different corrections may be associated with output maximization and input minimization. The difference can be substantial so this choice always deserves consideration. Furthermore, later in this book, we shall also introduce other models, where outputs are maximized and inputs are simultaneously minimized so that still further choices may need to be considered.

3.9 AN EXTENSION OF THE TWO PHASE PROCESS IN THE CCR MODEL

The two phase process described in Section 3.3 aims at obtaining the maximum sum of slacks (input excesses and output shortfalls). Hence it may occur that the projection of an inefficient DMU onto the efficient frontier is very far from the observed values in the sense that the input mix and output mix of the projected point are very different from those of the observations. It may be desirable to achieve the opposite approach: to project onto an efficient point which has input mix and output mix as near as possible to the observations. We propose an extension of the two phase process to achieve this.¹⁰ In this section we assume that the data set (X, Y) is positive, although the following process can be easily extended to the semipositive case.

First, we apply the two phase process to the data set and find the set R composed of CCR-efficient DMUs, i.e., DMUs with $\theta^* = 1$, $s^{-*} = 0$, $s^{+*} = 0$ for every optimal solution. We call R the “peer set.”

Then we go into Phase III process as follows. For each inefficient DMU_{*o*}, solve the following LP:

$$[\text{Phase III}] \quad \min w \quad (3.79)$$

$$\text{subject to } \theta^* x_{io} = \sum_{j \in R} x_{ij} \lambda_j + \delta_i^- x_{io} \quad (i = 1, \dots, m) \quad (3.80)$$

$$y_{ro} = \sum_{j \in R} y_{rj} \lambda_j - \delta_r^+ y_{ro} \quad (r = 1, \dots, s) \quad (3.81)$$

$$\delta_i^- \leq p \quad (i = 1, \dots, m) \quad (3.82)$$

$$\delta_r^+ \leq q \quad (r = 1, \dots, s) \quad (3.83)$$

$$\begin{aligned} p &\leq w, \quad q \leq w \\ \lambda_j &\geq 0, \quad \delta_i^- \geq 0, \quad \delta_r^+ \geq 0, \quad (\forall j, i, r) \end{aligned} \tag{3.84}$$

where θ^* is the optimal objective value obtained in Phase I.

The above LP can be interpreted as follows. The left side of (3.80) is the radially reduced value of input i of DMU_o, which is expressed as the sum of the weighted sum of inputs in the peer set and the slacks as designated by $\delta_i^- x_{io}$ on the right. Hence δ_i^- denotes the relative deviation ratio from the observed x_{io} . The inequality (3.82) bounds the maximum of δ_i^- ($i = 1, \dots, m$) by p . In the same way, we bound the maximum deviation ratio of output by q in (3.83). Finally, we put the bound w to p and q , and minimize w in the objective function (3.79). Therefore, we want to find slacks with the minimum deviation ratio from the observed values. (This is a kind of min-max problem.)

Let an optimal solution be $(\lambda^*, s^{-*}, s^{+*}, p^*, q^*, w^*)$. Then the following projection puts DMU_o on the efficient frontier.

$$\begin{aligned} \hat{x}_{io} &\leftarrow (\theta^* - \delta_i^{-*})x_{io} \quad (i = 1, \dots, m) \\ \hat{y}_{ro} &\leftarrow (1 + \delta_r^{+*})y_{ro} \quad (r = 1, \dots, s). \end{aligned}$$

We notice that this projection is “units invariant” in the sense that δ_i^{-*} and δ_r^{+*} exhibit no effects from a change of units for input i and output r , as can be seen from (3.80) and (3.81). A similar procedure is proposed in Coelli (1998).¹¹

[An Example]

We will apply the Phase III process to the following problem in Table 3.4 with 5 DMUs (A, B, C, D and E), 3 inputs (x_1, x_2 and x_3) and a single unitized output (y).

First we applied the CCR model with Phase I and Phase II processes and

Table 3.4. Problem for Phase III Process

DMU	x_1	x_2	x_3	y
A	1	2	1	1
B	1	1	2	1
C	2	10	5	1
D	2	5	10	1
E	2	10	10	1

obtained results displayed in Table 3.5. DMUs A and B are CCR-efficient and DMU A is the only member of the reference set used to evaluate inefficient DMUs C, D and E. For the input mix in Table 3.4, we find that C is similar to A rather than B so the use of A as a reference for C is reasonable. However,

Table 3.5. CCR-Score, Reference Set, Slacks and % Change

DMU	Score	Ref.	s_1	s_2	s_3	s_1/x_1	s_2/x_2	s_3/x_3
A	1	A	0	0	0	0%	0%	0%
B	1	B	0	0	0	0%	0%	0%
C	0.5	A	0	3	1.5	0%	30%	30%
D	0.5	A	0	0.5	4	0%	10%	40%
E	0.5	A	0	3	4	0%	30%	40%

D's input mix is similar to B rather than A and E's is likely to be similar to the combination of A and B. To accommodate these observations we applied the Phase III process as follows, using the peer set $R=\{A, B\}$ in the manner we now describe:

DMU C: The LP for C is as follows,

$$\begin{aligned}
 & \min \quad w \\
 \text{subject to} \quad & 0.5 \times 2 = 1 = \lambda_1 + \lambda_2 + 2\delta_1^- \\
 & 0.5 \times 10 = 5 = 2\lambda_1 + \lambda_2 + 10\delta_2^- \\
 & 0.5 \times 5 = 2.5 = \lambda_1 + 2\lambda_2 + 5\delta_3^- \\
 & 1 = \lambda_1 + \lambda_2 - \delta^+ \\
 & \delta_1^- \leq p, \delta_2^- \leq p, \delta_3^- \leq p, \delta^+ \leq q \\
 & p \leq w, q \leq w,
 \end{aligned}$$

where all variables are constrained to be nonnegative.

An optimal solution is given by

$$w^* = p^* = 0.3, \lambda_1^* = 1, \delta_2^{-*} = 0.3, \delta_3^{-*} = 0.3,$$

with all other variables zero. This solution is the same as the CCR solution, the maximum deviation ratio being 30% as in Table 3.5. The Phase III process projects C onto A.

DMU D: The LP for D is as follows,

$$\begin{aligned}
 & \min \quad w \\
 \text{subject to} \quad & 0.5 \times 2 = 1 = \lambda_1 + \lambda_2 + 2\delta_1^- \\
 & 0.5 \times 5 = 2.5 = 2\lambda_1 + \lambda_2 + 5\delta_2^- \\
 & 0.5 \times 10 = 5 = \lambda_1 + 2\lambda_2 + 10\delta_3^- \\
 & 1 = \lambda_1 + \lambda_2 - \delta^+ \\
 & \delta_1^- \leq p, \delta_2^- \leq p, \delta_3^- \leq p, \delta^+ \leq q \\
 & p \leq w, q \leq w,
 \end{aligned}$$

where all variables are constrained to be nonnegative.

An optimal solution is given by

$$w^* = p^* = 0.3, \lambda_2^* = 1, \delta_2^{-*} = 0.3, \delta_3^{-*} = 0.3,$$

with all other variables zero. This solution is different from the CCR solution. The Phase III process projects D onto B. The maximum deviation ratio 30% is less than that of the CCR model 40% which is given by the ratio of excess $s_3^{-*}/x_{3D} = 4/10 = 0.4(40\%)$.

DMU E: The LP for E is as follows,

$$\begin{aligned} \min \quad & w \\ \text{subject to} \quad & 0.5 \times 2 = 1 = \lambda_1 + \lambda_2 + 2\delta_1^- \\ & 0.5 \times 10 = 5 = 2\lambda_1 + \lambda_2 + 10\delta_2^- \\ & 0.5 \times 10 = 5 = \lambda_1 + 2\lambda_2 + 10\delta_3^- \\ & 1 = \lambda_1 + \lambda_2 - \delta^+ \\ & \delta_1^- \leq p, \delta_2^- \leq p, \delta_3^- \leq p, \delta^+ \leq q \\ & p \leq w, q \leq w, \end{aligned}$$

where all variables are constrained to be nonnegative.

An optimal solution is given by

$$w^* = p^* = 0.35, \lambda_1^* = 0.5, \lambda_2^* = 0.5, \delta_2^{-*} = 0.35, \delta_3^{-*} = 0.35,$$

with all other variables zero. This solution is different from the CCR solution. The Phase III process projects D onto the mid-point of A and B. The maximum deviation ratio 35% is less than that of the CCR model 40% which is given by the ratio of excess $s_3^{-*}/x_{3E} = 4/10 = 0.4(40\%)$.

The above results suggest that the Phase III process is suitable for finding a set of similar reference DMUs among the efficient DMUs, while keeping the CCR score at the level of the Phase I result.

3.10 DISCRETIONARY AND NON-DISCRETIONARY INPUTS

Up to this point we have assumed that all inputs and outputs can be varied at the discretion of management or other users. These may be called “discretionary variables.” “Non-discretionary variables,” not subject to management control, may also need to be considered. In evaluating performances of different bases for the Fighter Command of U.S. Air Forces, for instance, it was necessary to consider weather as an input since the number of “sorties” (successfully completed missions) and “aborts” (non-completed mission)¹² treated as outputs, could be affected by the weather (measured in degree days and numbers of “flyable” days) at different bases.

Even though Non-Discretionary, it is important to take account of such inputs in a manner that is reflected in the measures of efficiency used. We follow

the route provided by Banker and Morey (1986)¹³ who refer to such variables as “exogenously fixed,” in forms like “age of store,” in their use of DEA to evaluate the performances of 60 DMUs in a network of fast-food restaurants. Reverting to algebraic notation, we can represent their formulation by the following modification of the CCR model.

$$\min \quad \theta - \varepsilon \left(\sum_{i \in D} s_i^- + \sum_{r=1}^s s_r^+ \right) \quad (3.85)$$

$$\text{subject to} \quad \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in D \quad (3.86)$$

$$x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in ND \quad (3.87)$$

$$y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s. \quad (3.88)$$

where all variables (except θ) are constrained to be nonnegative.

Here the symbols $i \in D$ and $i \in ND$ refer to the sets of “Discretionary” and “Non-Discretionary” inputs, respectively. To be noted in the constraints is the fact that the variable θ is not applied to the latter inputs because these values are exogenously fixed and it is therefore not possible to vary them at the discretion of management. Therefore this is recognized by entering all x_{io} , $i \in ND$ at their fixed (observed) value. Turning to the objective (3.85) we utilize the symbol¹⁴ $\varepsilon > 0$ to mean that the slack variables (shown in the parentheses) are to be handled at a second stage where, as previously described, they are to be maximized in a manner that does not disturb the previously determined first-stage minimization of θ to achieve $\theta = \theta^*$. Finally, we note that the slacks s_i^- , $i \in ND$ are omitted from the objective. Hence these Non-Discretionary inputs do not enter directly into the efficiency measures being optimized in (3.85). They can, nevertheless, affect the efficiency evaluations by virtue of their presence in the constraints.

We can further clarify the way these Non-Discretionary variables affect the efficiency scores by writing the dual of (3.85)-(3.88) in the form of the following (modified) multiplier model,

$$\max \quad \sum_{r=1}^s u_r y_{ro} - \sum_{i \in ND} v_i x_{io} \quad (3.89)$$

$$\text{subject to} \quad \sum_{r=1}^s u_r y_{rj} - \sum_{i \in ND} v_i x_{ij} - \sum_{i \in D} v_i x_{ij} \leq 0, \quad j = 1, \dots, n \quad (3.90)$$

$$\sum_{i \in D} v_i x_{io} = 1 \quad (3.91)$$

$$v_i \geq \varepsilon, \quad i \in D \quad (3.92)$$

$$v_i \geq 0, \quad i \in ND \quad (3.93)$$

$$u_r \geq \varepsilon, \quad r = 1, \dots, s. \quad (3.94)$$

As can be seen, the Non-Discretionary but not the Discretionary inputs, enter into the objective (3.89). The multiplier values associated with these Non-Discretionary inputs may be zero, as in (3.93), but the other variables must always be positive, as in (3.92) and (3.94). The interpretations we now provide flow from the “complementary slackness principle” of linear programming. If (3.87) is satisfied strictly at an optimum then $v_i^* = 0$ is associated with this constraint and this x_{io} does not affect the evaluation recorded in (3.85). On the other hand, if $v_i^* > 0$ for any $i \in ND$ then the efficiency score recorded in (3.85) is reduced by the multiplier, x_{io} , for this DMU_o.

This follows from the dual theorem of linear programming—viz.,

$$\theta^* - \varepsilon \left(\sum_{i \in D} s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right) = \sum_{r=1}^s u_r^* y_{ro} - \sum_{i \in ND} v_i^* x_{io}. \quad (3.95)$$

Via this same relation we see that a *decrease* in this same x_{io} will *increase* the efficiency score recorded in the expression on the left of the equality (3.95).

For managerial use, the sense of this mathematical characterization may be interpreted in the following manner. The output achieved, as recorded in the y_{ro} , deserve a higher efficiency rating when they have been achieved under a relatively tighter constraint and a lower efficiency score when this constraint is loosened by increasing this x_{io} .

This treatment of Non-Discretionary variables must be qualified, at least to some extent, since, *inter alia*, allowance must be made for ranges over which the values of v_i^* and u_r^* remain unaltered. Modifications to allow for $v_i^* \geq \varepsilon \geq 0$ must also be introduced, as described in Problem 3.2 at the end of this chapter and effects like those treated in Problem 3.3 need to be recognized when achievement of the efficient frontier is a consideration.

After additional backgrounds have been supplied, Chapter 7, later in this book, introduces a variety of extensions and alternate ways of treating Non-Discretionary outputs as well as inputs. Here we only need to note that modifications of a relatively straightforward variety may also be made as in the following example.

A study of 638 public secondary schools in Texas was undertaken by a consortium of 3 universities in collaboration with the Texas Education Agency (TEA). The study was intended to try to develop improved methods for accountability and evaluation of school performances. In addition to discretionary inputs like teacher salaries, special instruction, etc., the following Non-Discretionary inputs had to be taken into account,

1. **Minority** : Number of minority students, expressed as a percentage of total student enrollment.
2. **Disadvantage** : Number of economically disadvantaged students, expressed as a percentage of total student enrollment.

3. LEP : Number of Limited English Proficiency students, expressed as a percentage of total student enrollment.

These inputs differ from the ones that Banker and Morey had in mind. For instance, a regression calculation made at an earlier stage of the consortium study yielded *negative* coefficients that tested statistically significant for every one of these 3 inputs in terms of their effects on academic test scores. In the subsequent DEA study it was therefore deemed desirable to reverse the sign associated with the x_{io} in the expression on the right of (3.95)

This was accomplished by reversing (3.87)

$$\text{from } \sum_{j=1}^n x_{ij} \lambda_j \leq x_{io} \text{ to } \sum_{j=1}^n x_{ij} \lambda_j \geq x_{io} \quad (3.96)$$

for each of these $i \in ND$. In this manner an ability to process more of these inputs, when critical, was recognized in the form of higher (rather than lower) efficiency scores. Full details are supplied in I.R. Bardhan (1995)¹⁵ and summarized in Arnold *et al.* (1997).¹⁶ Here we note that the results were sufficiently satisfactory to lead to a recommendation to accord recognition to “efficiency” as well as “effectiveness” in state budgetary allocations to different school districts. Although not accepted, the recommendation merits consideration at least to the extent of identifying shortcomings in the usual statements (and reward structures) for academic performance. The following scheme can help to clarify what is being said,¹⁷

Effectiveness implies

- Ability to state desired goals
- Ability to achieve desired goals

Efficiency relates to

- Benefits realized
- Resources used

Consider, for instance, the State-mandated Excellence Standards for Texas recorded in Table 3.6. These represent statements of desired goals and schools are rewarded (or not rewarded) on the basis of their achievements. Nothing is said about the amounts (or varieties) of resources used. Hence it should be no surprise that only 1 of the excellent-rated schools included in this study were found to be efficient. The others schools rated as excellent by the State of Texas had all expended excessive resources in achieving these goals. On the other hand many schools that failed to achieve excellence were nevertheless found to be efficient in producing desired outputs under very difficult conditions. Therefore some way of recognizing this kind of achievements is needed if “efficiency” as well as “effectiveness” is to be rewarded.¹⁸

Table 3.7 as taken from Bardhan (1994), contains portions of the printout from a DEA study involving one of these “excellent” schools. The row labelled

Table 3.6. State-mandated Excellence Standards on Student Outcomes

Outcome Indicator	State-mandated Excellence Standard
1. Texas Assessment of Academic Skills (TAAS) Test	90% of students passing on all standardized tests
2. Attendance	97% of total enrollment in the school
3. Dropout Rate	Less than or equal to 1% of total enrollment
4. Graduation Rate	99% of graduating class
5. College Admission Tests	<ul style="list-style-type: none"> • 35% of graduates scoring above the criterion score which is equal to 25 on the ACT^a and 1000 on the SAT^b • 70% of graduates taking either the ACT or the SAT

^aACT = American Collegiate Tests

^bSAT = Scholastic Aptitude Tests

Disadv (= Economically Disadvantaged) shows a slack value of 15 students. Judging from the performances of its peer group of efficient schools, this school

Table 3.7. Non-Discretionary Inputs

	Current Level	Slack	Value if Efficient
Minority	47.0	—	47.0
Disadv	14.0	15.0	29.0
LEP	3.5	—	3.5

should therefore have been able to more than double the number of such students that it processed *without affecting its efficiency score* because (a) the slack for $i \in ND$ is not in the objective of (3.85) and (b) the presence of nonzero slack for any $i \in ND$ means the associated $v_i^* = 0$. In addition, the other ND inputs are associated with positive multiplier values so this school should have been able to *increase* its efficiency score by *increasing* the Minority and LEP (Low English Proficiency = Mainly Hispanic) students it processed. As shown in Arnold *et al.* (1997) this can be done by introducing constraints to insure that no worsening of any already achieved excellence is avoided.

This is not the end of the line for what can be done. Additional extensions could proceed to a two-stage operation in which school “outputs” are transformed into “outcomes” where the latter includes things like success in securing employment *after* school has been completed.¹⁹ For instance, see C.A.K. Lovell *et al.* (1994)²⁰ who employ cohort data in such a two-stage analysis where

“outputs” are converted to “outcomes” to find that the record for American Public Schools is substantially better for the latter than the former.

3.11 SUMMARY OF CHAPTER 3

In this chapter we described the CCR model in some detail in both its input-oriented and output-oriented versions.

1. We also relaxed assumptions of a positive data set to semipositivity.
2. We defined the production possibility set based on the constant returns-to-scale assumption and developed the CCR model under this assumption.
3. The dual problem of the original CCR model was introduced as (DLP_o) in (3.6)-(3.9) and the existence of input excesses and output shortfalls clarified by solving this model. (To avoid confusion and to align our terminology with the DEA literature, we referred to the dual as the “envelopment model” and the primal (LP_o) introduced in Chapter 2 as the “multiplier model.”)
4. In Definition 3.2 we identified a DMU as *CCR-efficient* if and only if it is (i) radial-efficient and (ii) has zero-slack in the sense of Definition 3.1. Hence a DMU is CCR-efficient if and only if it has no input excesses and no output shortfalls.
5. Improvement of inefficient DMUs was discussed and formulae were given for effecting the improvements needed to achieve full CCR efficiency in the form of the CCR-projections given in (3.22) and (3.23).
6. Detailed computational procedures for the CCR model were presented in Section 3.6 and an optimal multiplier values (v , u) were obtained as the simplex multipliers for an optimal tableau obtained from the simplex method of linear programming.

3.12 NOTES AND SELECTED BIBLIOGRAPHY

As noted in Section 3.3 the term “Pareto-Koopmans” efficiency refers to Vilfredo Pareto and Tjalling Koopmans. The former, i.e., Pareto, was concerned with welfare economics which he visualized in terms of a vector-valued function with its components representing the utilities of all consumers. He then formulated the so-called Pareto condition of welfare maximization by noting that such a function could not be at a maximum if it was possible to increase one of its components without worsening other components of such a vector valued function. He therefore suggested this as a criterion for judging any proposed social policy: “the policy should be adopted if it made some individuals better off without decreasing the welfare (measured by their utilities) of other individuals.” This, of course, is a necessary but not a sufficient condition for maximizing such a function. Proceeding further, however, would involve comparing utilities to determine whether a decrease in the utilities of some persons would

be more than compensated by increases in the utilities of other individuals. Pareto, however, wanted to proceed as far as possible without requiring such interpersonal comparisons of utilities. See Vilfredo Pareto, *Manuel d'économie politique*, deuxième édition, Appendix, pp. 617 ff., Alfred Bonnet, ed. (Paris: Marcel Giard, 1927).

Tjalling Koopmans adapted these concepts to production. In an approach that he referred to as "activity analysis," Koopmans altered the test of a vector optimum by reference to whether it was possible to increase any output without worsening some other output under conditions allowed by available resources such as labor, capital, raw materials, etc. See pp. 33-97 in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation*, (New York: John Wiley & Sons, Inc., 1951).

The approaches by Pareto and Koopmans were entirely conceptual. No empirical applications were reported before the appearance of the 1957 article by M.J. Farrell in the *Journal of the Royal Statistical Society* under the title "The Measurement of Productive Efficiency." This article showed how these methods could be applied to data in order to arrive at relative efficiency evaluations. This, too, was in contrast with Koopmans and Pareto who conceptualized matters in terms of theoretically known efficient responses without much (if any) attention to how inefficiency, or more precisely, technical inefficiency, could be identified. Koopmans, for instance, assumed that producers would respond optimally to prices which he referred to as "efficiency prices." Pareto assumed that all consumers would (and could) maximize their utilities under the social policies being considered. The latter topic, i.e., the identification of inefficiencies, seems to have been, first, brought into view in an article published as "The Coefficient of Resource Utilization," by G. Debreu in *Econometrica* (1951) pp. 273-292. Even though their models took the form of linear programming problems both Debreu and Farrell formulated their models in the tradition of "activity analysis." Little, if any, attention had been paid to computational implementation in the activity analysis literature. Farrell, therefore, undertook a massive and onerous series of matrix inversions in his first efforts. The alternative of linear programming algorithms was called to Farrell's attention by A.J. Hoffman, who served as a commentator in this same issue of the *Journal of the Royal Statistical Society*. Indeed, the activity analysis approach had already been identified with linear programming and reformulated and extended in the 1957 article A. Charnes and W.W. Cooper published in "On the Theory and Computation of Delegation-Type Models: K-Efficiency, Functional Efficiency and Goals," *Proceedings of the Sixth International Meeting of The Institute of Management Science* (London: Pergamon Press). See also Chapter IX in A. Charnes and W.W. Cooper, *Management Models and Industrial Applications of Linear Programming* (New York: John Wiley & Sons, Inc., 1961).

The modern version of DEA originated in two articles by A. Charnes, W.W. Cooper and E. Rhodes: (1) "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, 1978, pp. 429-444 and (2)

"Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Management Science* 27, 1981, pp.668-697. The latter article not only introduced the name Data Envelopment Analysis for the concepts introduced in the former article, it also exploited the duality relations as well as the computational power that the former had made available. In addition, extensions were made that included the CCR projection operations associated with (3.32)-(3.33) and used these projections to evaluate programs (such as the U.S. Office of Education's "Program Follow Through") that made it possible to identify a "program's efficiency" separately from the way the programs had been managed. (Hence the distinction between the "program" and "managerial" efficiencies that had been confounded in the observations generated from this education study.) Equally important, the latter article started a tradition in which applications were used to guide subsequent research in developing new concepts and methods of analyses. This, in turn, led to new applications, and so on, with many hundreds now reported in the literature.

The first of the above two articles also introduced the ratio form of DEA that is represented as (FP_o) in Section 2.3 of Chapter 2. This ratio form with its enhanced interpretive power and its contacts with other definitions of efficiency in engineering and science was made possible by prior research in which Charnes and Cooper opened the field of fractional programming. See Problem 3.1, below. It also led to other new models and measures of efficiency which differ from (DLP_o) in Section 3.3, above, which had been the only model form that was previously used.

3.13 RELATED DEA-SOLVER MODELS FOR CHAPTER 3

CCR-I (Input-oriented Charnes-Cooper-Rhodes model).

This code solves the CCR model expressed by (3.6)-(3.9) or by (3.38)-(3.42). The data set should be prepared in an Excel Workbook by using an appropriate Workbook name prior to execution of this code. See the sample format displayed in Figure B.1 in Section B.5 of Appendix B and refer to explanations above the figure. This style is the most basic one and is adopted in other models as the standard main body of data. The main results will be obtained in the following Worksheets as displayed in Table 3.8. The worksheet "Summary" includes statistics on data — average, standard deviation of each input and output, and correlation coefficients between observed items. It also reports DMUs with inappropriate data for evaluation and summarizes the results.

CCR-O (Output-oriented CCR model).

This code solves the output-oriented CCR model expressed by (3.60)-(3.63). In this model the optimal efficiency score η^* describes the output enlargement rate and satisfies $\eta^* \geq 1$. However, we display this value by its inverse as $\theta^* = 1/\eta^*(\leq 1)$ and call it the "CCR-O efficiency score." This will facilitate comparisons of scores between the input-oriented and the output-

oriented models. In the CCR model, both models have the same efficiency value as shown by (3.64). The other results are exhibited in the Worksheets in Table 3.8. In Worksheet “Weight,” v and u correspond to p and q in (3.68)-(3.71), respectively. “Projection” is based on the formulas in (3.74) and (3.75).

Table 3.8. Worksheets Containing Main Results

Worksheet name	Contents
Summary	Summary on data and results.
Score	The efficiency score θ^* , the reference set(λ^*), ranking, etc.
Rank	The descending order ranking of efficiency scores.
Weight	The optimal (dual) multipliers v^* , u^* in (3.2)-(3.5).
WeightedData	The weighted data $\{x_{ij}v_i^*\}$ and $\{y_{rj}u_r^*\}$.
Slack	The input excesses s^- and the output shortfalls s^+ in (3.10).
Projection	Projection onto the efficient frontiers by (3.22)-(3.23).
Graph1	The bar chart of the CCR scores.
Graph2	The bar chart of scores in ascending order.

3.14 PROBLEM SUPPLEMENT FOR CHAPTER 3

Problem 3.1 (Ratio Forms and Strong and Weak Disposal)

The suggested response to Problem 2.3 in Chapter 2 showed how the “ratio of ratios” definition of efficiency in engineering could be subsumed under the CCR ratio form of DEA given for (FP_o) in Section 2.3 of that chapter. Can you now extend this to show how the linear programming formulation given in (LP_o) relates to the ratio form given in (FP_o) ?

Suggested Response : A proof of equivalence is provided with Theorem 2.1 in Section 2.4 of Chapter 2. The following proof is adopted from A. Charnes and W.W. Cooper, “Programming with Linear Fractional Functionals” *Naval Research Logistics Quarterly* 9, 1962, pp.181-185, which initiated (and named) the field of fractional programming. We use this as an alternative because it provides contact with the (now) extensive literature on fractional programming. To start we reproduce (FP_o) from section 2.3 of Chapter 2, as follows

$$\begin{aligned} \max \quad & \theta = \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \\ \text{subject to} \quad & \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \quad j = 1, \dots, n \\ & u_r, v_i \geq 0. \quad \forall r, i \end{aligned} \tag{3.97}$$

Now we can choose a new variable t in such a way that

$$t \sum_{i=1}^m v_i x_{io} = 1, \quad (3.98)$$

which implies $t > 0$. Multiplying all numerators and denominators by this t does not change the value of any ratio. Hence setting

$$\begin{aligned} \mu_r &= tu_r, \quad r = 1, \dots, s \\ \nu_i &= tv_i, \quad i = 1, \dots, m \end{aligned} \quad (3.99)$$

we have replaced the above problem by the following equivalent,

$$\begin{aligned} \max \quad \theta &= \sum_{r=1}^s \mu_r y_{ro} \\ \text{subject to} \quad \sum_{i=1}^m \nu_i x_{io} &= 1 \\ \sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m \nu_i x_{ij} &\leq 0 \quad j = 1, \dots, n \\ \mu_r, \nu_i &\geq 0. \quad \forall r, i \end{aligned} \quad (3.100)$$

This is the same as (LP_o) in Section 2.4 which we have transformed using what is referred to as the “Charnes-Cooper transformation” in fractional programming. This reduction of (3.97) to the linear programming equivalent in (3.100) also makes available (DLP_o) the dual problem which we reproduce here as

$$\begin{aligned} \min \quad \theta & \\ \text{subject to} \quad \theta x_{io} &= \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+. \quad r = 1, \dots, s \end{aligned} \quad (3.101)$$

This is the form used by Farrell (1978) which we employed for Phase I, as described in Section 3.3, followed by a Phase II in which the slacks are maximized in the following problem

$$\begin{aligned} \max \quad & \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \\ \text{subject to} \quad \theta^* x_{io} &= \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s \\ 0 &\leq \lambda_j, s_i^-, s_r^+, \quad \forall j, i, r \end{aligned} \quad (3.102)$$

where θ^* is the value obtained by solving (3.101) in Phase I. As noted in the text the solution to (3.101) is referred to as “Farrell efficiency.” It is also referred to as “weak efficiency,” as measured by θ^* , since this measure does not comprehend the non-zero slacks that may be present. In the economics literature (3.101) is said to assume “strong disposal.” If we omit the s_i^- in the first m constraints in (3.101) we then have what is called “weak disposal” which we can write in the following form

$$\begin{aligned} \min \quad & \theta \\ \text{subject to} \quad & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j, \quad i = 1, \dots, m \\ & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s \end{aligned} \tag{3.103}$$

which means that the input inequalities are replaced with equalities so there is no possibility of positive input slacks that may have to be disposed of. Sometimes this is referred to as the assumption of “weak” and “strong” *input* disposal in order to distinguish it from corresponding formulations in *output* oriented models. In either case, these weak and strong disposal assumptions represent refinements of the “free disposal” assumption introduced by T.C. Koopmans (1951) for use in activity analysis. This assumption means that there is no cost associated with disposing of excess slacks in inputs or outputs. That is, slacks in the objective are all to be assigned a zero coefficient. Hence, all nonzero slacks are to be ignored, whether they occur in inputs or outputs. For a fuller treatment of “weak” and “strong” disposal, see R. Färe, S. Grosskopf and C.A.K. Lovell, *The Measurement of Efficiency of Production* (Boston: Kluwer Academic Publishers Group, 1985).

Problem 3.2

Can you provide a mathematical formulation that will serve to unify the Phase I and Phase II procedures in a single model?

Suggested Response : One way to do this is to join (3.6) and (3.11) together in a single objective as follows:

$$\begin{aligned} \min \quad & \theta - \varepsilon \left(\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{subject to} \quad & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i = 1, \dots, m \\ & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s \\ & 0 \leq \lambda_j, s_i^-, s_r^+. \quad \forall j, i, r \end{aligned} \tag{3.104}$$

It is tempting to represent $\varepsilon > 0$ by a small (real) number such as $\varepsilon = 10^{-6}$. However, this is not advisable. It can lead to erroneous results and the situation may be *worsened* by replacing $\varepsilon = 10^{-6}$ by even smaller values. See I. Ali and L. Seiford, "The Mathematical Programming Approach to Efficiency Analysis" in H.O. Fried, C.A.K. Lovell and S.S. Schmidt, ed., *The Measurement of Productive Efficiency* (New York: Oxford University Press, 1993) or Ali and Seiford (1993) "Computational Accuracy and Infinitesimals in Data Envelopment Analysis," *INFOR* 31, pp.290-297.

As formulated in the Charnes, Cooper and Rhodes article in the *European Journal of Operational Research*, cited in the Notes and Selected Bibliography, above, $\varepsilon > 0$ is formulated as a "non-Archimedean infinitesimal." That is, $\varepsilon > 0$ is smaller than *any* positive real number and, in fact, the product of ε by *any* real number, so that, however large the multiplier, $k > 0$, the value of $k\varepsilon > 0$ remains smaller than *any* positive real number. This means that $\varepsilon > 0$ is not a real number because the latter all have the Archimedean property — viz., given any real number $n > 0$ there exists another real number $n/2$ such that $n > n/2 > 0$. Thus, to deal with non-Archimedean elements it is necessary to embed the field of real numbers in a still larger field. However, it is not necessary to go into the further treatments of this kind of (non-standard) mathematics. It is not even necessary to specify a value of $\varepsilon > 0$ explicitly. The two-phase procedure described in this chapter accomplishes all that is required. Phase I accords priority to $\min \theta = \theta^*$ with $\theta^* > 0$ when the data are semipositive. Fixing $\theta = \theta^*$ as is done in (3.12) for Phase II in Section 3.3 of this chapter, we must then have, by definition,

$$0 < \theta^* - \varepsilon \left(\sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right).$$

A more detailed treatment of these non-Archimedean elements and their relations to mathematical programming may be found in V. Arnold, I. Bardhan, W.W. Cooper and A. Gallegos, "Primal and Dual Optimality in Computer Codes Using Two-Stage Solution Procedures in DEA" in J.E. Aronson and S. Zionts, ed., *Operations Research: Methods, Models and Applications* (Westport, Conn.: Quorum Books, 1998). Here we only note that the further implications of this non-Archimedean element, which appears in the objective of (3.104) can be brought forth by writing its dual as

$$\begin{aligned} \max \quad & \theta = \sum_{r=1}^s \mu_r y_{ro} \\ \text{subject to} \quad & \sum_{r=1}^s \mu_r y_{rj} - \sum_{i=1}^m \nu_i x_{ij} \leq 0 \quad j = 1, \dots, n \\ & \sum_{i=1}^m \nu_i x_{io} = 1 \\ & \mu_r, \nu_i \geq \varepsilon > 0. \quad \forall r, i \end{aligned} \tag{3.105}$$

This means that all variables are constrained to positive values. Hence, they are to be accorded “some” worth, even though it is not specified explicitly. Finally, as shown in Arnold *et al.*, the principle of complementary slackness used in Section 3.3 of this chapter is modified to the following

$$0 \leq s_i^{-*} \nu_i^* \leq s_i^{-*} \varepsilon \quad (3.106)$$

$$0 \leq s_r^{+*} \mu_r^* \leq s_r^{+*} \varepsilon. \quad (3.107)$$

Hence, unlike what is done with free disposal, one cannot justify ignoring nonzero slacks by assuming that corresponding multiplier values will be zero. Indeed, the stage two optimization maximizes the slacks, as is done in the Phase II procedure associated with (3.102), in order to try to wring out the maximum possible inefficiency values associated with nonzero slacks.

Problem 3.3

We might note that (3.105), above, differs from (3.100) by its inclusion of the non-Archimedean conditions $\mu_r, \nu_i \geq \varepsilon > 0$. Can you show how these conditions should be reflected in a similarly altered version of the ratio model in (3.97)?

Suggested Answer: Multiply and divide the objective in (3.105) by $t > 0$. Then multiply all constraints by the same t to obtain

$$\begin{aligned} & \max \frac{\sum_{r=1}^s (t\mu_r)y_{ro}}{t} && (3.108) \\ \text{subject to } & \sum_{r=1}^s (t\mu_r)y_{rj} - \sum_{i=1}^m (t\nu_i)x_{ij} \leq 0 \quad j = 1, \dots, n \\ & \sum_{i=1}^m (t\nu_i x_{io}) = t \\ & (t\mu_r), (t\nu_i) \geq \varepsilon > 0. \quad \forall r, i \end{aligned}$$

Set $u_r = t\mu_r$, $v_i = t\nu_i$ for each r and i . Then substitute in these last expressions to obtain

$$\begin{aligned} & \max \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} && (3.109) \\ \text{subject to } & \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \quad j = 1, \dots, n \\ & u_r / \sum_{i=1}^m v_i x_{io} \geq \varepsilon \quad \forall r \\ & v_i / \sum_{i=1}^m v_i x_{io} \geq \varepsilon. \quad \forall i \end{aligned}$$

Note that we have reflected the condition $\sum_{i=1}^m v_i x_{io} = t$ in the above objective but not in the constraints because it can *always* be satisfied. We can also now

write

$$\frac{\sum_{r=1}^s u_r^* y_{ro}}{\sum_{i=1}^m v_i^* x_{io}} = \sum_{r=1}^s u_r^* y_{ro} = \theta^* - \varepsilon \left(\sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*} \right). \quad (3.110)$$

The equality on the right follows from the dual theorem of linear programming. We have added the equality on the left by virtue of our derivation and, of course, the values of these expressions are bounded by zero and one.

Problem 3.4

The optimal solution for (LP_o) , the multiplier problem, for DMU A in Example 3.1 is

$$v_1^* = 0.1429, v_2^* = 0.1429, u^* = 0.8571.$$

Show how these multiplier values could be used as a management guide to improve efficiency in A 's performance.

Suggested Response : Applying the above solution to the data in Table 3.2 we utilize the multiplier model to write the constraints for DMU A as follows:

DMU	Output Value	Output Variable	Input Variable	Input Value
A	0.8571	$= 1u^*$	$\leq 4v_1^* + 3v_2^*$	$= 1.000$
B	0.8571	$= 1u^*$	$\leq 7v_1^* + 3v_2^*$	$= 1.429$
C	0.8571	$= 1u^*$	$\leq 8v_1^* + v_2^*$	$= 1.284$
$\rightarrow D$	0.8571	$= 1u^*$	$\leq 4v_1^* + 2v_2^*$	$= 0.8571$
$\rightarrow E$	0.8571	$= 1u^*$	$\leq 2v_1^* + 4v_2^*$	$= 0.8571$
F	0.8571	$= 1u^*$	$\leq 10v_1^* + v_2^*$	$= 1.5719$
G	0.8571	$= 1u^*$	$\leq 3v_1^* + 7v_2^*$	$= 1.429$

$$\text{Add constraint : } 1 = 4v_1^* + v_2^*.$$

Note, first, that $1u^* = \theta^* = 0.8571$ so the value of the optimal solution of the multiplier model is equal to the optimal value of the envelopment model in accordance with the duality theory of linear programming and, similarly, the reference set is $\{D, E\}$, as indicated by the arrows. Here we have $v_1^* = v_2^* = 0.1429$ so a unit reduction in the $4+3=7$ units of input used by A would bring it into $4+2=6$ units used by D and E . Thus, $v^* = v_1^* = v_2^* = 0.1429 = 1/7$ is the reduction required to bring A to full efficiency. We can provide added perspective by reducing $x_1 = 4$ to $\hat{x}_1 = 3$ which positions A half way between D and E on the efficient frontier in Figure 3.2. It also produces equality of the relations for A as in the following expression

$$u^* = 3v_1^* + 3v_2^* \doteq 0.8571.$$

Division then produces

$$\frac{u^*}{3v_1^* + 3v_2^*} = 1,$$

which is the condition for efficiency prescribed for (FP_o) as given in (2.3)-(2.6) in Chapter 2. Here we have $v_1^* = v_2^* = 0.1429$ so a use of either x_1 or x_2

produces equal results per unit change en route to achieving full efficiency. More generally we will have $v_1^* \neq v_2^*$, etc., so these values may be used to guide managerial priorities for the adjustments to the model.

There are two exceptions that require attention. One exception is provided by Definition 2.1 in Chapter 2, which notes that all components of the vectors \mathbf{v}^* and \mathbf{u}^* must be positive in at least one optimal solution. We use the following solution for G as an illustration,

$$v_1^* = 0.3333, \quad v_2^* = 0, \quad u^* = 0.6666.$$

From the data of Example 3.1, we then have

$$0.6666 = 1u^* < 3v_1^* + 7v_2^* = 1.00.$$

Replacing $x_1 = 3$ by $\hat{x}_1 = 2$ produces equality, but efficiency cannot be claimed because $v_2^* = 0$ is present.

Reference to Figure 3.2 shows that the inefficiency associated with the nonzero slack in going from R to E is not attended to and, indeed, the amount of slack associated with x_2 is worsened by the horizontal movement associated with the reduction in going from $x_1 = 3$ to $\hat{x}_1 = 2$. Such a worsening would not occur for F in Figure 3.2, but in either case full efficiency associated with achieving a ratio of unity does not suffice unless all multipliers are positive in the associated adjustments.

The second exception noted above involves the evaluations of DMUs which are fully efficient and hence are used to evaluate *other* DMUs. This complication is reflected in the fact that the data in the basis matrix B is changed by any such adjustment and so the inverse B^{-1} used to evaluate other DMUs, as given in (3.53), is also changed. This topic involves complexities which cannot be treated here, but is attended to later in this book after the requisite background has been supplied. It is an important topic that enters into further uses of DEA and hence has been the subject of much research which we summarize as follows.

The problem of sensitivity to data variations was first addressed in the DEA literature by A. Charnes, W.W. Cooper, A.Y. Lewin, R.C. Morey and J. Rousseau "Sensitivity and Stability Analysis in DEA," *Annals of Operations Research* 2, 1985, pp.139-156. This paper was restricted to sensitivity analyses involving changes in a single input or output for an efficient DMU. Using the CCR ratio model, as in (3.97), above, this was subsequently generalized to simultaneous changes in all inputs and outputs for any DMU in A. Charnes and L. Neralic, "Sensitivity Analysis of the Proportionate Change of Inputs (or Outputs) in Data Envelopment Analysis," *Glasnik Matematicki* 27, 1992, pp.393-405. Also using the ratio form of the multiplier model, R.G. Thompson and R.M. Thrall provided sensitivity analyses in which *all* data for *all* DMUs are varied simultaneously. The basic idea is as follows. All inputs and all outputs are worsened for the efficient DMUs until at least one DMU changes its status. The multipliers and ratio values are then recalculated and the process is reiterated until all DMUs become inefficient. See R.G. Thompson, P.S. Dharmapala, J. Diaz, M. Gonzalez-Lima and R.M. Thrall, "DEA Multiplier Analytic Center

Sensitivity Analysis with an Illustrative Application to Independent Oil Companies," *Annals of Operations Research* 66, 1996, pp.163-167. An alternate and more exact approach is given in L. Seiford and J. Zhu, "Sensitivity Analysis of DEA Models for Simultaneous Changes in All the Data," *Journal of the Operational Research Society* 49, 1998, pp.1060-1071. Finally, L. Seiford and J. Zhu in "Stability Regions for Maintaining Efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 108, 1998, pp. 127-139 develop a procedure for determining exact stability regions within which the efficiency of a DMU remains unchanged. More detail on the topic of sensitivity is covered in Chapter 9. See also W.W. Cooper, L.M. Seiford and J. Zhu (2004) Chapter 3, "Sensitivity Analysis in DEA," in W.W. Cooper, L.M. Seiford and J. Zhu, eds., *Handbook on Data Envelopment Analysis* (Norwell, Mass., Kluwer Academic Publishers).

Problem 3.5 (Complementary Slackness and Sensitivity Analysis)

Table 3.9 displays a data set of 5 stores (A, B, C, D and E) with two inputs (the number of employees and the floor area) and two outputs (the volume of sales and profits).

Assignment : (i) Using the code CCR-I in the DEA-Solver, obtain the efficiency score (θ^*), reference set (λ^*), the optimal weights (v^*, u^*) and slacks (s^{-*}, s^{+*}) for each store. (ii) Interpret the complementary slackness conditions (3.15) between the optimal weights and the optimal slacks. (iii) Discuss the meaning of u_1^* of C , and use sensitivity analysis to verify your discussion. (iv) Check the output shortfall s_2^{+*} of B and identify the influence of its increase on the value of the corresponding multiplier (weight) u_2 using the CCR-I code. (v) Regarding the store B , discuss the meaning of v_1^* .

Suggested Response : (i) The efficiency score (θ^*) and reference set (λ^*) are listed in Table 3.9 under the heading "Score." Store E is the only efficient

Table 3.9. Data and Scores of 5 Stores

Store	Data				θ^*	Score Rank	Reference
	Employee	Area	Sales	Profits			
A	10	20	70	6	0.933333	2	$E (0.77778)$
B	15	15	100	3	0.888889	3	$E (1.11111)$
C	20	30	80	5	0.533333	5	$E (0.88889)$
D	25	15	100	2	0.666667	4	$E (1.11111)$
E	12	9	90	8	1	1	$E (1)$

DMU and is referent to all other stores. The optimal weights and slacks are displayed in Table 3.10.

Table 3.10. Optimal Weights and Slacks

Store	Weights					Slacks			
	Emply. v_1^*	Area v_2^*	Sales u_1^*	Profits u_2^*	Emply. s_1^{-*}	Area s_2^{-*}	Sales s_1^{+*}	Profits s_2^{+*}	
A	0.1	0	0.0133	0	0	11.6667	0	0.2222	
B	0.0667	0	0.00889	0	0	3.3333	0	5.8889	
C	0.05	0	0.00667	0	0	8	0	2.1111	
D	0	0.0667	0.00667	0	3.3333	0	0	6.8889	
E	0.0702	0.0175	0.00526	0.0658	0	0	0	0	

(ii) The complementary slackness conditions in (3.15) assert that, for an optimal solution we have

$$v_i^* s_i^{-*} = 0 \text{ (for } i = 1, 2\text{)} \text{ and } u_r^* s_r^{+*} = 0 \text{ (for } r = 1, 2\text{).}$$

This means that if $v_i^* > 0$ then $s_i^{-*} = 0$ and if $s_i^{-*} > 0$ then $v_i^* = 0$, and the same relations hold between u^* and s^{+*} . This can be interpreted thus: if a DMU has an excess (s^{-*}) in an input against the referent DMU, then the input item has no value in use to DMU_o being evaluated, so the optimal (multiplier) solution assigns a value of zero to the corresponding weight (v^*). Similarly we can see that for a positive weight to be assigned to a multiplier, the corresponding optimal slack must be zero.

(iii) C has $u_1^* = 0.00667$ and this value can be interpreted in two ways. First, with recourse to the fractional program described by (2.3)-(2.6) in Chapter 2, we can interpret u_1^* (the optimal weight to “sales”) as the degree of influence which one unit of sales has on the optimal efficiency score. Thus, if C increases its sales by one unit, then it is expected that θ^* will increase by 0.00667 ($=u_1^*$) to $0.533333 + 0.00667 = 0.54$, since the denominator of (2.3) does not change and retains the value 1. On the other hand, $u_2^* = 0$ implies that a one unit increase in “profits” has no effect on the efficiency score of C. The second interpretation is that in (3.59) we introduced u^* as “pricing vectors” and in the envelopment form of LP, u_r^* is the reduced cost induced by one unit change in output r . Notice that y_o appears as the constant in (3.41). Hence $u_1^* = 0.00667$ shows the degree of contribution that one unit change of sales can make to use the efficiency score. The above observations contribute to identifying which output has the largest influence on the efficiency score. However, such analysis of influence is valid only in a limited range of observations. To check this range you can change the efficiency score of C by adding to the data set a virtual store C', with sales of 101 and other data equal to C, and then applying the CCR-I code again.

(iv) B has the slack $s_2^{+*} = 5.889$ (shortfall in “profits”) and, by the complementary slackness theorem, $u_2^{+*} = 0$, showing that this level of profits has no

effect on the efficiency evaluation. Adding this value to current profits will raise profits to 8.889. We augmented the data set by adding B' , which has profits of 8.889, and other observations equal to B and tried CCR-I for this data set. As a result, we have, for B' , $s_2^{+*} = 0$ and $u_2^{+*} = 0.1$. This means that, at this level of profits, this output has a positive influence on the efficiency score, which in this case is a 0.1 per unit increase.

(v) B has the optimal weight $v_1^* = 0.0667$. It can therefore be said that changes in the number of employees affects the efficiency score. In this case, reduced costs analysis using LP is not as straightforward as in u^* . Hence, dealing with the fractional program in (2.3)-(2.6) is appropriate for the sensitivity analysis of data on employees. If B reduces its employees by one to 14, then the denominator of (2.3) decreases by $v_1^* = 0.0667$. Since the other terms are unchanged, the objective function (2.3) takes the value

$$(0.00889 \times 100 + 0 \times 3) / (0.0667 \times 14 + 0 \times 15) = 0.952.$$

Since the value is still less than one and all constraints in (2.4) are satisfied by this set of (v^*, u^*) , the above objective value gives a lower bound of θ^* for this updated problem. Application of CCR-I code for this new problem showed that 0.952 is the optimal objective value. Thus, it may be expected that an input item with a large v_i^* value greatly affects the efficiency value.

Problem 3.6

Solve Example 3.1 in Table 3.2 (Section 3.7) using the output-oriented model (CCR-O model in DEA-Solver) and compare the corresponding CCR-projection with the input-oriented case.

Suggested Response : For ease of comparison with the input-oriented case, the optimal objective value η^* is demonstrated by its inverse θ^* ($= 1/\eta^*$).

The optimal solution to A reads

$$\begin{aligned}\theta^* &= 0.8571 \quad (\eta^* = 1.1667) \\ \lambda_D^* &= 0.8333, \quad \lambda_E^* = 0.3333, \quad \text{other } \lambda_j^* = 0 \\ s_1^{-*} &= s_2^{-*} = s^{+*} = 0\end{aligned}$$

The output-oriented CCR-projection was performed following (3.74)-(3.75), which resulted in

$$\begin{aligned}\hat{x}_1 &\leftarrow x_1 = 4 \quad (\text{no change}) \\ \hat{x}_2 &\leftarrow x_2 = 3 \quad (\text{no change}) \\ \hat{y} &\leftarrow \eta^* y = 1.167 \times 1 = 1.167 \quad (16.7\% \text{ increase}).\end{aligned}$$

These results differ from the input-oriented case, reflecting the difference in model characteristics. Table 3.11 exhibits the CCR-projection in both input and output orientations.

Table 3.11. CCR-projection in Input and Output Orientations

DMU	Data			Score θ^*	Input orientation			Output orientation		
	x_1	x_2	y		x_1	x_2	y	x_1	x_2	y
A	4	3	1	0.8571	3.43	2.57	1	4	3	1.17
B	7	3	1	0.6316	4.42	1.89	1	7	3	1.58
C	8	1	1	1.0000	8	1	1	8	2	1
D	4	2	1	1.0000	4	2	1	4	2	1
E	2	4	1	1.0000	2	4	1	2	4	1
F	10	1	1	1.0000	8	1	1	8	1	1
G	3	7	1	0.6667	2	4	1	3	6	1.5

Problem 3.7

Suppose the activity (x_o, y_o) is inefficient. Let the improved activity obtained by the input-oriented model (3.22) and (3.23) be (\hat{x}_o, \hat{y}_o) ,

$$\hat{x}_o = \theta^* x_o - s^{-*} = X\lambda^* \quad (3.111)$$

$$\hat{y}_o = y_o + s^{+*} = Y\lambda^*, \quad (3.112)$$

while that of the output-oriented model (3.74) and (3.75) be $(\check{x}_o, \check{y}_o)$.

$$\check{x}_o = x_o - t^{-*} = X\mu^* \quad (3.113)$$

$$\check{y}_o = \eta^* y_o + t^{+*} = Y\mu^*. \quad (3.114)$$

An activity on the line segment connecting these two points can be expressed by

$$(\dot{x}_o, \dot{y}_o) = \alpha_1 (\hat{x}_o, \hat{y}_o) + \alpha_2 (\check{x}_o, \check{y}_o) \quad (3.115)$$

$$\alpha_1 + \alpha_2 = 1, \quad \alpha_1 \geq 0, \quad \alpha_2 \geq 0.$$

Prove the following

Proposition 3.1 *The activity (\dot{x}_o, \dot{y}_o) is CCR-efficient.*

Suggested Answer :

Proof. Since by (3.65) $\mu^* = \lambda^*/\theta^*$, we have

$$\begin{aligned} (\dot{x}_o, \dot{y}_o) &= \alpha_1 (X, Y)\lambda^* + \alpha_2 (X, Y)\lambda^*/\theta^* \\ &= (\alpha_1 + \alpha_2/\theta^*) (X\lambda^*, Y\lambda^*) \end{aligned}$$

By applying Theorem 3.4, we can see that (\dot{x}_o, \dot{y}_o) is CCR-efficient. \square

This proposition is valid for *any* nonnegative combination of two points ($\alpha_1 \geq 0$ and $\alpha_2 \geq 0$). That is, it holds even when relaxing the convex-combination condition $\alpha_1 + \alpha_2 = 1$. Also, the case $\alpha_1 = 1, \alpha_2 = 0$ corresponds to the input-oriented improvement and the case $\alpha_1 = 0, \alpha_2 = 1$ to the output-oriented one. The case $\alpha_1 = 1/2, \alpha_2 = 1/2$ is a compromise between the two points.

Problem 3.8

Solve the hospital example in Chapter 1 (see Table 1.5) by input-oriented and output-oriented CCR models. Compare the CCR-projections of hospital C in both cases. Based on the previous Proposition 3.1, determine the combination of both projections using (3.115) with $\alpha_1 = \alpha_2 = 1/2$.

Suggested Answer : The data sheet for this problem is recorded in Figure B.1 in Appendix B. Prepare two Excel Workbooks, e.g., “Hospital-CCR-I.xls” and “Hospital-CCR-O.xls,” each containing Figure B.1 as the data sheet. Then run CCR-I (the input-oriented CCR model) and CCR-O (the output-oriented CCR model). After the computations, the CCR-projections are stored in “Projection” sheets as follows:

For input orientation:

Doctor	25	→	20.9	(16% reduction)
Nurse	160	→	141	(12% reduction)
Outpatient	160	→	160	(no change)
Inpatient	55	→	55	(no change)

For output orientation:

Doctor	25	→	23.6	(5% reduction)
Nurse	160	→	160	(no change)
Outpatient	160	→	181	(13% increase)
Inpatient	55	→	62.3	(13% increase)

We have a compromise of the two models as the averages of the projected values as follows (this is equivalent to the case $\alpha_1 = \alpha_2 = 1/2$ in Proposition 3.1):

Doctor	25	→	22	(12% reduction)
Nurse	160	→	150	(6% reduction)
Outpatient	160	→	171	(7% increase)
Inpatient	55	→	59	(7% increase)

These improvements will put C on the efficient frontier.

Problem 3.9

Assume that the data set (X, Y) is semipositive (see the definition in Section 3.2) and let an optimal solution of (DLP_o) in Section 3.3 be $(\theta^*, \lambda^*, s^{-*}, s^{+*})$. Prove that the reference set defined by $E_o = \{j | \lambda^* > 0\}$ ($j \in \{1, \dots, n\}$) is not empty.

Suggested Answer :

We have the equation:

$$\theta^* x_o = X \lambda^* + s^{-*}.$$

Let an optimal solution of (LP_o) in Section 3.3 be (v^*, u^*) . By multiplying v^* from left to the above equation, we have

$$\theta^* v^* x_o = v^* X \lambda^* + v^* s^{-*}.$$

Since $\mathbf{v}^* \mathbf{x}_o = 1$ as a constraint in (LP_o) and $\mathbf{v}^* \mathbf{s}^{-*} = 0$ by the complementarity condition, it holds

$$\theta^* = \mathbf{v}^* \mathbf{X} \boldsymbol{\lambda}^*.$$

Here, θ^* is positive as claimed in Section 3.3. Thus, $\boldsymbol{\lambda}^*$ must be semipositive and E_o is not empty. \square

Problem 3.10

Part 1. Using model (3.101) show that the output inequality will always be satisfied as an equation in the single output case. That is, the output slack will be zero at an optimum.

Proof :

It is easy to see what is happening if we start with the case of no outputs. In this case the solution will be $\theta_o^* = 0$ and all $\lambda_j^* = 0$ since only the output constraints keep this from happening. It follows that for the case of one output a solution with $y_o < \sum_{j=1}^n y_j \lambda_j$ cannot be optimal since this choice of λ s would prevent θ from achieving its minimum value. To see that this is so note that a choice of the minimizing value of θ is determined by

$$\theta = \max_i \left\{ \frac{\sum_{j=1}^n x_{ij} \lambda_j}{x_{io}} \mid i = 1, \dots, m \right\} = \frac{\sum_{j=1}^n x_{ij} \lambda_j}{x_{io}}$$

and this maximum value can be lowered until $y_o = \sum_{j=1}^n y_j \lambda_j$. (Here for simplicity we are assuming that all data are positive.) Hence optimality requires $y_o = \sum_{j=1}^n y_j \lambda_j^*$ so the output slack is zero in an optimum solution. \square

Part 2. Extend the above to show that at least one output inequality must be satisfied as an equation in the case of multiple outputs $r = 1, \dots, s$, like the ones represented in (3.101).

Proof :

Replace (3.101) by the following model

$$\min_{\boldsymbol{\lambda}, \theta} \max_r \left\{ \theta, \sum_{j=1}^n y_{rj} \lambda_j \geq y_{ro} \mid r = 1, \dots, s \right\} = (\theta^*, \sum_{j=1}^n y_{kj} \lambda_j^* = y_{ko})$$

subject to (3.116)

$$\theta x_{[io]} \geq \sum_{j=1}^n x_{ij} \lambda_j, \quad i = 1, \dots, m$$

$$0 \leq \theta \leq 1, \quad \lambda_j \geq 0, \quad j = 1, \dots, n.$$

This is a multiple criteria programming problem which seeks to minimize the maximum of the $r = 1, \dots, s$ inequalities representing the output values set by the y_{ro} as lower limits for each of the s outputs. Note that the maximum output values are limited by the input constraints. The condition $0 \leq \theta \leq 1$ which eliminates the possibility of infinite solutions is not needed in (3.101) since its minimizing objective guarantees its fulfillment.

Now the maximal values may be decreased in (3.116) until equality is achieved in at least one of the output inequalities and the minimization of θ eliminates the possibility of being misled by the choice of a set of λ values that does not minimize this maximum because of the presence of alternate optima. The optimizing values θ^* , λ^* represented on the right of (3.116) therefore minimize θ with

$$\theta^* = \max_i \left\{ \frac{\sum_{j=1}^n x_{ij} \lambda_j^*}{x_{io}}, i = 1, \dots, m \right\}$$

and all constraints are satisfied in (3.101) as well as in (3.116). \square

Corollary 3.2 *At least one input as well as one output constraint will be satisfied as an equality at an optimum so these constraints have zero slack in (3.101) and (3.116).*

We also have following

Theorem 3.6 *Model (3.116) is equivalent to model (3.101).*

Similar results hold for the “output-oriented” version of the CCR model with a “Farrell” measure of efficiency that is represented by the value of θ^* in (3.101). Such equivalences may also be established for other DEA models, like the BCC model that will be presented in the next chapter. Finally, we also note that we earlier established the relation of (3.101) to a fractional programming model so our linear programming DEA formulation provides a link between this multiple objective nonlinear programming problem and a nonlinear nonconvex fractional programming problem. See also Cooper (2005)²¹ for a discussion of relations between multiple criteria programming and goal programming formulations which are also equivalent to a nonlinear problem directed to minimizing a sum of absolute values.

Notes

1. It might be noted that Postulate (A2) is included in (A4) but is separated out for special attention.
2. See Appendix A.4.
3. See Appendix A.4.
4. Farrell also restricted his treatments to the single output case. See M.J. Farrell (1957), “The Measurement of Production Efficiency,” *Journal of the Royal Statistical Society A*, 120, pp.253-281.
5. In the linear programming literature this is called the “complementary slackness” condition. This terminology is given by A.W. Tucker who is responsible for formulating and proving this. See E.D. Nering and A.W. Tucker *Linear Programs and Related Problems*, (Harcourt Brace, 1993). See Appendix A.6 for a detailed development.
6. See Appendix A.8.
7. Although empirical studies show the uniqueness of the reference set for most DMUs, there may be multiple reference sets and improvement plans (projections) in the existence of multiple optimal solutions.
8. See Appendix A.4.

9. We notice that OQ and OA are measured by some “distance measure.” If we employ the “Euclidian measure” – also called the “ l_2 metric” – we have

$$\frac{d(OQ)}{d(OA)} = \frac{\sqrt{3.428^2 + 2.571^2}}{\sqrt{4^2 + 3^2}} = \frac{\sqrt{18.36}}{\sqrt{25}} = 8.57.$$

However, the measure is not restricted to Euclidian measure. Any l_k measure gives the same result provided $k < \infty$. See Appendix A in Charnes and Cooper, *Management Models and Industrial Applications of Linear Programming* (New York, John Wiley, Inc., 1961).

10. K. Tone “An Extension of the Two Phase Process in the CCR Model,” in *Proceedings of the 1999 Fall National Conference of the Operations Research Society of Japan*.

11. T. Coelli (1998), “A Multi-stage Methodology for the Solution of Oriented DEA Models,” *Operations Research Letters* 22, pp.143-149.

12. “Aborts” were treated as reciprocals so that an increase in their output would reduce the value of the numerator in the (FP_o) objective represented in (2.3). They could also have been subtracted from a dominantly large positive constant. See A. Charnes, T. Clark, W.W. Cooper and B. Golany “A Development Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Forces,” *Annals of Operational Research* 2, 1985, pp.59-94.

13. R.D. Banker and R.C. Morey (1986), “Efficiency Analysis for Exogenously Fixed Inputs and Outputs,” *Operations Research* 34, 1986, pp.513-521. See also Chapter 10 in R. Färe, S. Grosskopf and C.A. Knox Lovell *Production Frontiers* (Cambridge University Press, 1994) where this is referred to as “sub-vector optimizations.”

14. A fuller treatment of these $\epsilon > 0$ values is provided in Problem 3.2 at the end of this chapter.

15. I.R. Bardhan (1995), “DEA and Stochastic Frontier Regression Approaches Applied to Evaluating Performances of Public Secondary Schools in Texas,” Ph.D. Thesis. Austin Texas: Graduate School of Business, the University of Texas at Austin. Also available from University Microfilms, Inc. in Ann Arbor, Michigan.

16. V.L. Arnold, I.R. Bardhan and W.W. Cooper “A Two-Stage DEA Approach for Identifying and Rewarding Efficiency in Texas Secondary Schools” in W.W. Cooper, S. Thore, D. Gibson and F. Phillips, eds., IMPACT: *How IC² Research Affects Public Policy and Business Practices* (Westport, Conn.: Quorum Books, 1997)

17. An additional category may also be identified as *Property* which applies to (a) Choice of goals or objectives, and (b) Choice of means to achieve goals or objectives.

18. See Arnold *et al.* (1997) for further suggestions and discussions.

19. See Governmental Accounting Standards Board (GASB) Research Report: “Service Efforts and Accomplishments Reporting: Its Time Has Come,” H.P. Hatry, J.M. Sullivan, J.M. Fountain and L. Kremer, eds. (Norwell, Conn., 1990).

20. C.A.K. Lovell, L.C. Walters and Lisa Wood, “Stratified Models of Education Production Using Modified DEA and Regression Analysis,” in A. Charnes, W.W. Cooper, A.Y. Lewin and L.M. Seiford, eds., *Data Envelopment Analysis: Theory, Methodology and Applications* (Norwell, Mass.: Kluwer Academic Publishers, 1994).

21. W.W. Cooper (2005), “Origin, Uses of and Relations Between Goal Programming and DEA (Data Envelopment Analysis),” *Journal of Multiple Criteria Decision Analysis* (to appear).

4 ALTERNATIVE DEA MODELS

4.1 INTRODUCTION

In the preceding chapters, we discussed the CCR model, which is built on the assumption of *constant* returns to scale of activities as depicted for the production frontier in the single input-single output case shown in Figure 4.1. More generally, it is assumed that the production possibility set has the following property: If (\mathbf{x}, \mathbf{y}) is a feasible point, then $(t\mathbf{x}, t\mathbf{y})$ for any positive t is also feasible. This assumption can be modified to allow production possibility sets with different postulates. In fact, since the very beginning of DEA studies, various extensions of the CCR model have been proposed, among which the BCC (Banker-Charnes-Cooper) model¹ is representative. The BCC model has its production frontiers spanned by the convex hull of the existing DMUs. The frontiers have piecewise linear and concave characteristics which, as shown in Figure 4.2, leads to *variable* returns-to-scale characterizations with (a) increasing returns-to-scale occurring in the first solid line segment followed by (b) decreasing returns-to-scale in the second segment and (c) constant returns-to-scale occurring at the point where the transition from the first to the second segment is made.

In this chapter, we first introduce the BCC model in Section 2. Then, in Section 3, the “Additive Model” will be described. This model has the same production possibility set as the BCC and CCR models and their variants

but treats the slacks (the input excesses and output shortfalls) directly in the objective function.

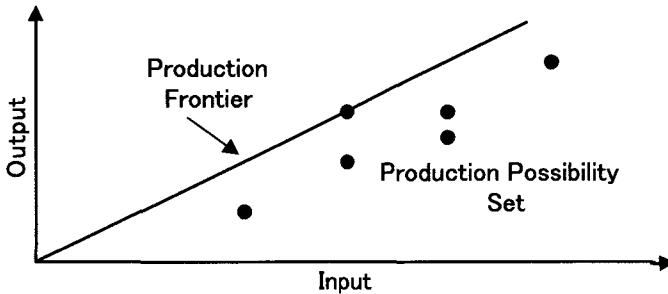


Figure 4.1. Production Frontier of the CCR Model

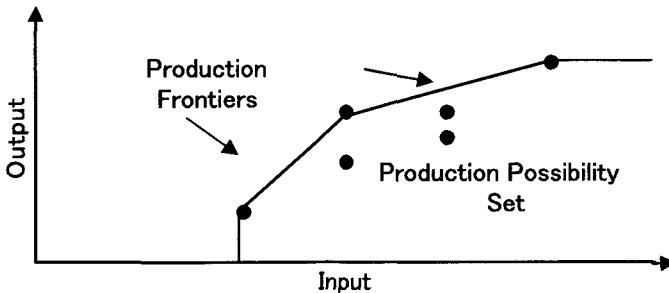


Figure 4.2. Production Frontiers of the BCC Model

CCR-type models, under weak efficiency, evaluate the radial (proportional) efficiency θ^* but do not take account of the input excesses and output shortfalls that are represented by non-zero slacks. This is a drawback because θ^* does not include the nonzero slacks. Although the Additive model deals with the input excesses and output shortfalls directly and can discriminate efficient and inefficient DMUs, it has no means to gauge the depth of inefficiency by a scalar measure similar to the θ^* in the CCR-type models. To make up for this deficiency, we will introduce a slacks-based measure of efficiency (SBM), which was proposed by Tone (1997)² and is also referred to as the "Enhanced Russell Measure," in Section 4 in the form of a scalar with a value not greater than the corresponding CCR-type measure θ^* . This measure reflects nonzero slacks in inputs and outputs when they are present.

The BCC and CCR models differ only in that the former, but not the latter, includes the convexity condition $\sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, \forall j$ in its constraints. Thus, as might be expected, they share properties in common and exhibit differences. They also share properties with the corresponding Additive models. Thus, the Additive model without the convexity constraint will characterize a DMU as efficient if and only if the CCR model characterizes it as efficient. Similarly, the BCC model will characterize a DMU as efficient if and only if the corresponding Additive model also characterizes it as efficient.

There are two types of measures or approaches in DEA: *radial* and *non-radial*. Differences exist in the characterization of input or output items. Suppose that there are four inputs x_1, x_2, x_3 and x_4 , in the problem of interest, where x_1 and x_2 are radial, and x_3 and x_4 are non-radial. That is, (x_1, x_2) are subject to change proportionally, such as $(\alpha x_1, \alpha x_2)$ (with $\alpha > 0$), while x_3 and x_4 are subject to change non-radially. These differences should be reflected in the evaluation of efficiency. Analogously, the output part can be divided into the radial and non-radial outputs. The radial approach is represented by the CCR and BCC models. Its shortcoming is that it neglects the non-radial input/output slacks. The non-radial approach SBM deals with slacks directly, but it neglects the radial characteristics of inputs and/or outputs. In Section 5, we integrate these approaches in a unified framework and present a hybrid measure of efficiency (Hybrid).

The concept of “translation invariance” that will be introduced in this chapter deals with lateral shifts of the constraints so that negative data, for instance, may be converted to positive values that admit of treatment by our solution methods, which assume that all data are non-negative. As we will see, the Additive models which include the convexity constraint, are translation invariant but this is not true when the convexity constraint is omitted. CCR models are also not translation invariant while BCC models are translation invariant to changes in the data for only some of their constraints.

These (and other) topics treated in this chapter will provide an overview of model selection possibilities and this will be followed up in further detail in chapters that follow.

The CCR, BCC, Additive, SBM and Hybrid models do not exhaust the available DEA models. Hence in the Appendix to this chapter we introduce a variant, called the Free Disposal Hull (FDH) model which assumes a nonconvex (staircase) production possibility set. Finally, in the problems we ask readers to use their knowledge of the Additive model to develop yet another, known as the “multiplicative model.”

4.2 THE BCC MODELS

Let us begin this section with a simple example. Figure 4.3 exhibits 4 DMUs, A, B, C and D , each with one input and one output.

The efficient frontier of the CCR model is the dotted line that passes through B from the origin. The frontiers of the BCC model consists of the bold lines connecting A, B and C . The production possibility set is the area consisting

of the frontier together with observed or possible activities with an excess of input and/or a shortfall in output compared with the frontiers. A , B and C are on the frontiers and BCC-efficient. The same is true for all points on the solid lines connecting A and B , and B and C . However, only B is CCR-efficient.

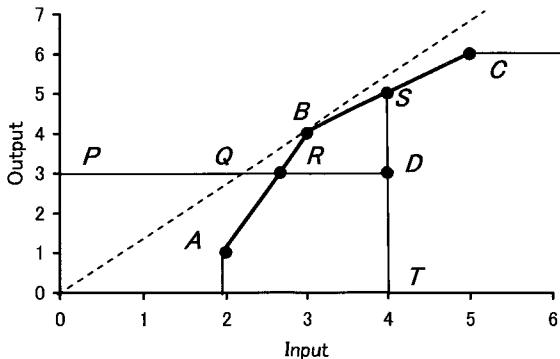


Figure 4.3. The BCC Model

Reading values from the this graph, the BCC-efficiency of D is evaluated by

$$\frac{PR}{PD} = \frac{2.6667}{4} = 0.6667,$$

while its CCR-efficiency is smaller with value

$$\frac{PQ}{PD} = \frac{2.25}{4} = 0.5625.$$

Generally, the CCR-efficiency does not exceed BCC-efficiency.

In the output-oriented BCC model, we read from the vertical axis of Figure 4.3 to find D evaluated by

$$\frac{ST}{DT} = \frac{5}{3} = 1.6667.$$

This means that achievement of efficiency would require augmenting D 's output from its observed value to $1.6667 \times 3 = 5$ units. The comparable augmentation for the CCR model is obtained from the reciprocal of its input inefficiency — viz., $1/0.5625 = 1.7778$ so, as the diagram makes clear, a still greater augmentation is needed to achieve efficiency. (Note: this simple “reciprocal relation” between input and output efficiencies is not available for the BCC model.)

Banker, Charnes and Cooper (1984) published the BCC model whose production possibility set P_B is defined by:

$$P_B = \left\{ (\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \mathbf{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0} \right\}, \quad (4.1)$$

where $X = (\mathbf{x}_j) \in R^{m \times n}$ and $Y = (\mathbf{y}_j) \in R^{s \times n}$ are a given data set, $\boldsymbol{\lambda} \in R^n$ and \mathbf{e} is a row vector with all elements equal to 1. The BCC model differs from

the CCR model only in the adjunction of the condition $\sum_{j=1}^n \lambda_j = 1$ which we also write $e\lambda = 1$ where e is a row vector with all elements unity and λ is a column vector with all elements non-negative. Together with the condition $\lambda_j \geq 0$, for all j , this imposes a convexity condition on allowable ways in which the observations for the n DMUs may be combined.

4.2.1 The BCC Model

The input-oriented BCC model evaluates the efficiency of DMU _{o} ($o = 1, \dots, n$) by solving the following (envelopment form) linear program:

$$(BCC_o) \quad \min_{\theta_B, \lambda} \theta_B \quad (4.2)$$

$$\text{subject to } \theta_B x_o - X \lambda \geq 0 \quad (4.3)$$

$$Y \lambda \geq y_o \quad (4.4)$$

$$e \lambda = 1 \quad (4.5)$$

$$\lambda \geq 0, \quad (4.6)$$

where θ_B is a scalar.

The dual multiplier form of this linear program (BCC_o) is expressed as:

$$\max_{v, u, u_0} z = u y_o - u_0 \quad (4.7)$$

$$\text{subject to } v x_o = 1 \quad (4.8)$$

$$-v X + u Y - u_0 e \leq 0 \quad (4.9)$$

$$v \geq 0, \quad u \geq 0, \quad u_0 \text{ free in sign}, \quad (4.10)$$

where v and u are vectors and z and u_0 are scalars and the latter, being “free in sign,” may be positive or negative (or zero). The equivalent BCC fractional program is obtained from the dual program as:

$$\max \frac{u y_o - u_0}{v x_o} \quad (4.11)$$

$$\text{subject to } \frac{u y_j - u_0}{v x_j} \leq 1 \quad (j = 1, \dots, n) \quad (4.12)$$

$$v \geq 0, \quad u \geq 0, \quad u_0 \text{ free}. \quad (4.13)$$

Correspondences between the primal-dual constraints and variables can be diagrammed as in Table 4.1.

It is clear that a difference between the CCR and BCC models is present in the free variable u_0 , which is the dual variable associated with the constraint $e\lambda = 1$ in the envelopment model that also does not appear in the CCR model.

The primal problem (BCC_o) is solved using a two-phase procedure similar to the CCR case. In the first phase, we minimize θ_B and, in the second phase, we maximize the sum of the input excesses and output shortfalls, keeping $\theta_B = \theta_B^*$ (the optimal objective value obtained in Phase one). The evaluations secured

Table 4.1. Primal and Dual Correspondences in BCC Model

Linear Programming Form			
Envelopment form constraints	Multiplier form variables		Envelopment form variables
$\theta_B \mathbf{x}_o - X\boldsymbol{\lambda} \geq \mathbf{0}$	$\mathbf{v} \geq \mathbf{0}$		$\mathbf{v}\mathbf{x}_o = 1$
$Y\boldsymbol{\lambda} \geq \mathbf{y}_o$	$\mathbf{u} \geq \mathbf{0}$		$-\mathbf{v}X + \mathbf{u}Y - u_0 \mathbf{e} \leq \mathbf{0}$
$e\boldsymbol{\lambda} = 1$	u_0		$\boldsymbol{\lambda} \geq \mathbf{0}$

from the CCR and BCC models are also related to each other as follows. An optimal solution for (BCC_o) is represented by $(\theta_B^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$, where \mathbf{s}^{-*} and \mathbf{s}^{+*} represent the maximal input excesses and output shortfalls, respectively. Notice that θ_B^* is not less than the optimal objective value θ^* of the CCR model, since (BCC_o) imposes one additional constraint, $e\boldsymbol{\lambda} = 1$, so its feasible region is a subset of feasible region for the CCR model.

Definition 4.1 (BCC-Efficiency)

If an optimal solution $(\theta_B^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ obtained in this two-phase process for (BCC_o) satisfies $\theta_B^* = 1$ and has no slack ($\mathbf{s}^{-*} = \mathbf{0}$, $\mathbf{s}^{+*} = \mathbf{0}$), then the DMU_o is called BCC-efficient, otherwise it is BCC-inefficient.

Definition 4.2 (Reference Set)

For a BCC-inefficient DMU_o , we define its reference set, E_o , based on an optimal solution $\boldsymbol{\lambda}^*$ by

$$E_o = \{ j \mid \lambda_j^* > 0 \} \quad (j \in \{1, \dots, n\}). \quad (4.14)$$

If there are multiple optimal solutions, we can choose any one to find that

$$\theta_B^* \mathbf{x}_o = \sum_{j \in E_o} \lambda_j^* \mathbf{x}_j + \mathbf{s}^{-*} \quad (4.15)$$

$$\mathbf{y}_o = \sum_{j \in E_o} \lambda_j^* \mathbf{y}_j - \mathbf{s}^{+*}. \quad (4.16)$$

Thus, we have a formula for improvement via the *BCC-projection*,

$$\hat{\mathbf{x}}_o \Leftarrow \theta_B^* \mathbf{x}_o - \mathbf{s}^{-*} \quad (4.17)$$

$$\hat{\mathbf{y}}_o \Leftarrow \mathbf{y}_o + \mathbf{s}^{+*}. \quad (4.18)$$

The following two theorems and the lemma can be proved in a way similar to ones used for the CCR model. Thus, analogous to Theorem 3.2 for the CCR model, we have

Theorem 4.1 *The improved activity (\hat{x}_o, \hat{y}_o) is BCC-efficient.*

Similarly, the following adapts Lemma 3.1 to the BCC model,

Lemma 4.1 *For the improved activity (\hat{x}_o, \hat{y}_o) , there exists an optimal solution $(\hat{v}_o, \hat{u}_o, \hat{u}_0)$ for its dual problem such that*

$$\hat{v}_o > \mathbf{0} \text{ and } \hat{u}_o > \mathbf{0} \quad (4.19)$$

$$\hat{v}_o \mathbf{x}_j = \hat{u}_o \mathbf{y}_j - \hat{u}_0 \quad (j \in E_o) \quad (4.20)$$

$$\hat{v}_o X \geq \hat{u}_o Y - \hat{u}_0 \mathbf{e}. \quad (4.21)$$

Theorem 4.2 *Every DMU in E_o associated with a $\lambda_j^* > 0$ as defined in (4.14), is BCC-efficient.*

This is an extension of Theorem 3.3. Finally, however, the next theorem exposes a property of BCC-efficiency for the input-oriented version of the model. This property is not secured by the CCR model, so the two may be used to check whether this property is present.

Theorem 4.3 *A DMU that has a minimum input value for any input item, or a maximum output value for any output item, is BCC-efficient.*

Proof. Suppose that DMU_o has a minimum input value for input 1, i.e., $x_{1o} < x_{1j}$ ($\forall j \neq o$). Then, from (4.3) and (4.5), DMU_o has the unique solution ($\theta_B^* = 1, \lambda_o^* = 1, \lambda_j^* = 0$ ($\forall j \neq o$)). Hence, DMU_o has $\theta_B^* = 1$ with no slacks and is BCC-efficient. The maximum output case can be proved analogously. \square

4.2.2 The Output-oriented BCC Model

Turning to the output-oriented BCC model we write

$$(BCC - O_o) \quad \max_{\eta_B, \boldsymbol{\lambda}} \eta_B \quad (4.22)$$

$$\text{subject to} \quad X \boldsymbol{\lambda} \leq \mathbf{x}_o \quad (4.23)$$

$$\eta_B \mathbf{y}_o - Y \boldsymbol{\lambda} \leq \mathbf{0} \quad (4.24)$$

$$\mathbf{e} \boldsymbol{\lambda} = 1 \quad (4.25)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \quad (4.26)$$

This is the envelopment form of the output-oriented BCC model. The dual (multiplier) form associated with the above linear program $(BCC - O_o)$ is expressed as:

$$\min_{\mathbf{v}, \mathbf{u}, v_0} z = \mathbf{v} \mathbf{x}_o - v_0 \quad (4.27)$$

$$\text{subject to} \quad \mathbf{u} \mathbf{y}_o = 1 \quad (4.28)$$

$$\mathbf{v} X - \mathbf{u} Y - v_0 \mathbf{e} \geq \mathbf{0} \quad (4.29)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, v_0 \text{ free in sign}, \quad (4.30)$$

where v_0 is the scalar associated with $e\lambda = 1$ in the envelopment model. Finally, we have the equivalent (BCC) fractional programming formulation for the latter (multiplier) model:

$$\min \frac{vx_o - v_0}{uy_o} \quad (4.31)$$

$$\text{subject to } \frac{vx_j - v_0}{uy_j} \geq 1 \quad (j = 1, \dots, n) \quad (4.32)$$

$$v \geq 0, \quad u \geq 0, \quad v_0 \text{ free in sign.} \quad (4.33)$$

4.3 THE ADDITIVE MODEL

The preceding models required us to distinguish between input-oriented and output-oriented models. Now, however, we combine both orientations in a single model, called the *Additive model*.

4.3.1 The Basic Additive Model

There are several types of Additive models, from which we select following:

$$(ADD_o) \quad \max_{\lambda, s^-, s^+} z = es^- + es^+ \quad (4.34)$$

$$\text{subject to } X\lambda + s^- = x_o \quad (4.35)$$

$$Y\lambda - s^+ = y_o \quad (4.36)$$

$$e\lambda = 1 \quad (4.37)$$

$$\lambda \geq 0, \quad s^- \geq 0, \quad s^+ \geq 0. \quad (4.38)$$

A variant, which we do not explore here, is an Additive model³ which omits the condition $e\lambda = 1$.

The dual problem to the above can be expressed as follows:

$$\min_{v, u, u_0} w = vx_o - uy_o + u_0 \quad (4.39)$$

$$\text{subject to } vX - uY + u_0 e \geq 0 \quad (4.40)$$

$$v \geq e \quad (4.41)$$

$$u \geq e \quad (4.42)$$

$$u_0 \text{ free.} \quad (4.43)$$

To explain this model we use Figure 4.4, where four DMUs A , B , C and D , each with one input and one output, are depicted. Since, by (4.35)–(4.38), the model (ADD_o) has the same production possibility set as the BCC model, the efficient frontier, which is continuous, consists of the line segments \overline{AB} and \overline{BC} . Now consider how DMU D might be evaluated. A feasible replacement of D with s^- and s^+ is denoted by the arrows s^- and s^+ in the figure. As shown by the dotted line in the figure, the maximal value of $s^- + s^+$ is attained at B . It is clear that this model considers the input excess and the output

shortfall simultaneously in arriving at a point on the efficient frontier which is most distant from D .⁴

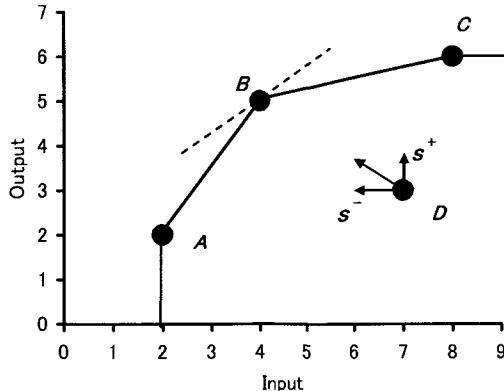


Figure 4.4. The Additive Model

Taking these considerations into account we can obtain a definition of efficiency as follows for the Additive model.

Let the optimal solutions be $(\lambda^*, s^{-*}, s^{+*})$. The definition of efficiency for an efficient DMU in the Additive model is then given by:

Definition 4.3 (ADD-efficient DMU)

DMU_o is ADD-efficient if and only if $s^{-*} = 0$ and $s^{+*} = 0$.

Theorem 4.4 DMU_o is ADD-efficient if and only if it is BCC-efficient.

A proof of this theorem may be found in Ahn *et al.*⁵ Here, however, it suffices to note that the efficiency score θ^* is not measured explicitly but is implicitly present in the slacks s^{-*} and s^{+*} . Moreover, whereas θ^* reflects only Farrell (=weak) efficiency, the objective in (ADD_o) reflects all inefficiencies that the model can identify in *both* inputs and outputs.

Theorem 4.5 Let us define $\hat{x}_o = x_o - s^{-*}$ and $\hat{y}_o = y_o + s^{+*}$. Then, (\hat{x}_o, \hat{y}_o) is ADD-efficient.

By this theorem, improvement to an efficient activity is attained by the following formulae (Projection for the Additive model):

$$\hat{x}_o \Leftarrow x_o - s^{-*} \quad (4.44)$$

$$\hat{y}_o \Leftarrow y_o + s^{+*}, \quad (4.45)$$

with (\hat{x}_o, \hat{y}_o) serving as the coordinates for the point on the efficient frontier used to evaluate DMU_o .

Example 4.1

We clarify the above with Table 4.2 which shows 8 DMUs with one input and one output. The solution of the Additive model and that of the BCC model are both exhibited. The reference set is determined as the set of DMUs which are in the optimal basic set of the LP problem. It is thus of interest to note that B , C and E are all fully efficient under both the BCC and (ADD_o) model, as asserted in Theorem 4.4. Furthermore, the nonzero slack under s^{+*} for A shows that it is not fully efficient for the BCC model (as well as (ADD_o)) even though it is weakly efficient with a value of $\theta^* = 1$. This means that A is on a part of the frontier that is not efficient, of course, while all of the other values with $\theta^* < 1$ mean that they fail even to achieve the frontier and, as can be observed, they also have nonzero slacks in their optimum (2^{nd} -stage) solutions. For the latter observations, therefore, a projection to $\hat{\mathbf{x}}_o = \theta^* \mathbf{x}_o$ need not achieve efficiency. It could merely bring the thus adjusted performance onto a portion of the frontier which is not efficient — as was the case for A .

Table 4.2. Data and Results of Example 4.1

DMU	Input		Output	BCC θ^*	Additive Model		
	x	y			s^{-*}	s^{+*}	Ref.
A	2	1	1	1	0	1	C
B	3	3	1	1	0	0	B
C	2	2	1	1	0	0	C
D	4	3	0.75	0.75	1	0	B
E	6	5	1	1	0	0	E
F	5	2	0.40	0.40	3	0	C
G	6	3	0.50	0.50	3	0	B
H	8	5	0.75	0.75	2	0	E

The following definition brings up another important distinction between the Additive model and BCC (or CCR) models;

Definition 4.4 (Mix) *We define “Mix” as proportions in which inputs are used or in which outputs are produced.*

Returning to the BCC (or CCR) models, it can be observed that $(1 - \theta^*)$ represents reductions which can be achieved without altering the input mix utilized and $(\eta^* - 1)$ would play a similar role for output expansions which do not alter the output mix.

In the literature of DEA, as well as economics, this proportion is referred to as “technical inefficiency” as contrasted with “scale inefficiencies,” “allocative (or price) inefficiencies” and other type of inefficiencies that will be discussed in subsequent chapters.

With all such technical inefficiencies accounted for in the BCC and CCR models, the second stage optimization is directed to maximize the slacks in order to see whether further inefficiencies are present. Altering any nonzero slack obtained in the second stage optimization must then necessarily alter the mix.

As is now apparent, the CCR and BCC models distinguish between technical and mix inefficiencies and this is a source of trouble in trying to arrange for a single measure of efficiency. The allocative model makes no such distinction, however, and so in Section 4.4 we will use the opportunity provided by these models to develop a measure that comprehends *all* of the inefficiencies that the model can identify.

As noted above, there is a distinction to be made between the technical and mix inefficiencies identified by the BCC and CCR models. The DEA literature (much of it focused on weak efficiency) does not make this distinction and we will follow the literature in its rather loose usage of “technical efficiency” to cover both types of inefficiency. When it is important to do so, we should refine this terminology by using “purely technical inefficiency” for θ^* and η^* and distinguish this from the mix inefficiencies given in Definition 4.4.

4.3.2 Translation Invariance of the Additive Model

In many applications it may be necessary (or convenient) to be able to handle negative data in some of the inputs or outputs. For instance, in order to determine whether mutual insurance companies are more (or less) efficient than their stock-ownership counterparts, it was necessary to be able to go beyond the assumption of semipositive data (as defined at the start of Chapter 3) in order to handle losses as well as profits treated as output.⁶ This was dealt with by employing a property of (ADD_o) known as “translation invariance” which we now develop from the following definition.

Definition 4.5 (Translation Invariance)

Given any problem, a DEA model is said to be translation invariant if translating the original input and/or output data values results in a new problem that has the same optimal solution for the envelopment form as the old one.

First we examine the input-oriented BCC model. In Figure 4.5, DMU D has the BCC-efficiency PR/PD . This ratio is invariant even if we shift the output value by changing the origin from O to O' . Thus, the BCC model is translation invariant with respect to *outputs* (but not *inputs*). Similar reasoning shows that the output oriented BCC model is invariant under the translation of *inputs* (but not *outputs*). (See Problem 4.3.)

Turning to the Additive model, Figure 4.6 shows that this model is translation invariant in both *inputs* and *outputs*, since the efficiency evaluation does not depend on the origin of the coordinate system when this model is used.

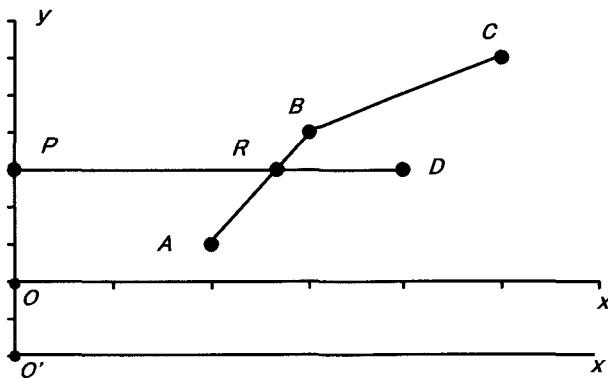


Figure 4.5. Translation in the BCC Model

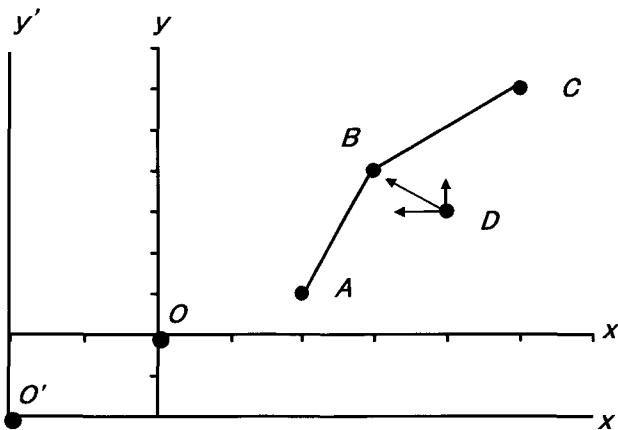


Figure 4.6. Translation in the Additive Model

We now develop this property of the Additive model in detail.

Let us translate the data set (X, Y) by introducing arbitrary constants $(\alpha_i : i = 1, \dots, m)$ and $(\beta_r : r = 1, \dots, s)$ to obtain new data

$$x'_{ij} = x_{ij} + \alpha_i \quad (i = 1, \dots, m : j = 1, \dots, n) \quad (4.46)$$

$$y'_{rj} = y_{rj} + \beta_r, \quad (r = 1, \dots, s : j = 1, \dots, n) \quad (4.47)$$

To show that this model is invariant under this arbitrary translation we observe that the x values (4.35) become

$$\sum_{j=1}^n (x'_{ij} - \alpha_i) \lambda_j + s_i^- = \sum_{j=1}^n x'_{ij} \lambda_j + s_i^- - \alpha_i = x'_{io} - \alpha_i$$

so that

$$\sum_{j=1}^n x'_{ij} \lambda_j + s_i^- = x'_{io} \quad (i = 1, \dots, m)$$

which are the same λ_j, s_i^- that satisfy

$$\sum_{j=1}^n x_{ij} \lambda_j + s_i^- = x_{io}. \quad (i = 1, \dots, m)$$

Similarly, the same λ_j, s_r^+ that satisfy

$$\sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{ro}. \quad (r = 1, \dots, s)$$

will also satify

$$\sum_{j=1}^n y'_{rj} \lambda_j - s_r^+ = y'_{ro}. \quad (r = 1, \dots, s)$$

Notice that the convexity condition $e\lambda = 1$ is a key factor in deriving the above relations. The above equalities show that if $(\lambda^*, s^{-*}, s^{+*})$ is an optimal solution of the original problem, then it is also optimal for the translated problem and *vice versa*. Finally, we also have

$$\begin{aligned} s_i^- &= \sum_{j=1}^n x_{ij} \lambda_j - x_{io} = \sum_{j=1}^n x'_{ij} \lambda_j - x'_{io} \\ s_r^+ &= \sum_{j=1}^n y_{rj} \lambda_j - y_{ro} = \sum_{j=1}^n y'_{rj} \lambda_j - y'_{ro} \end{aligned}$$

so the value of the objective is also not affected and we therefore have the following theorem for the Additive model (4.34)-(4.38) from Ali and Seiford (1990).⁷

Theorem 4.6 (Ali and Seiford (1990)) *The Additive model given by (4.34)-(4.38) is translation invariant.*

4.4 A SLACKS-BASED MEASURE OF EFFICIENCY (SBM)

We now augment the Additive models by introducing a measure that makes its efficiency evaluation, as effected in the objective, invariant to the units of measure used for the different inputs and outputs. That is, we would like this summary measure to assume the form of a scalar that yields the same efficiency value when distances are measured in either kilometers or miles. More generally, we want this measure to be the same when x_{io} and x_{ij} are replaced by $k_i x_{io} = \hat{x}_{io}$, $k_i x_{ij} = \hat{x}_{ij}$ and y_{ro} and y_{rj} are replaced by $c_r y_{ro} = \hat{y}_{ro}$, $c_r y_{rj} = \hat{y}_{rj}$ where the k_i and c_r are arbitrary positive constants, $i = 1, \dots, m$; $r = 1, \dots, s$.

This property is known by names such as “dimension free”⁸ and “units invariant.” In this section we introduce such a measure for Additive models in the form of a single scalar called “SBM,” (Slacks-Based Measure) which was introduced by Tone (1997, 2001) and has the following important properties:

1. (P1) The measure is invariant with respect to the unit of measurement of each input and output item. (Units invariant)
2. (P2) The measure is monotone decreasing in each input and output slack. (Monotone)

4.4.1 Definition of SBM

In order to estimate the efficiency of a DMU $(\mathbf{x}_o, \mathbf{y}_o)$, we formulate the following fractional program in λ , s^- and s^+ .

$$(SBM) \quad \min_{\lambda, s^-, s^+} \rho = \frac{1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{io}}{1 + \frac{1}{s} \sum_{r=1}^s s_r^+ / y_{ro}} \quad (4.48)$$

subject to $\mathbf{x}_o = X\lambda + s^-$
 $\mathbf{y}_o = Y\lambda - s^+$
 $\lambda \geq 0, s^- \geq 0, s^+ \geq 0.$

In this model, we assume that $X \geq O$. If $x_{io} = 0$, then we delete the term s_i^- / x_{io} in the objective function. If $y_{ro} \leq 0$, then we replace it by a very small positive number so that the term s_r^+ / y_{ro} plays a role of penalty.

It is readily verified that the objective function value ρ satisfies (P1) because the numerator and denominator are measured in the same units for every item in the objective of (4.48). It is also readily verified that an increase in either s_i^- or s_r^+ , all else held constant, will decrease this objective value and, indeed, do so in a strictly monotone manner.

Furthermore, we have

$$0 \leq \rho \leq 1. \quad (4.49)$$

To see that this relation holds we first observe that $s_i^- \leq x_{io}$ for every i so that $0 \leq s_i^- / x_{io} \leq 1$ ($i = 1, \dots, m$) with $s_i^- / x_{io} = 1$ only if the evidence shows that only a zero amount of this input was required. It follows that

$$0 \leq \frac{\sum_{i=1}^m s_i^- / x_{io}}{m} \leq 1.$$

This same relation does not hold for outputs since an output shortfall represented by a nonzero slack can exceed the corresponding amount of output produced. In any case, however, we have

$$0 \leq \frac{\sum_{r=1}^s s_r^+ / y_{ro}}{s}.$$

Thus these represent ratios of average input and output mix inefficiencies with the upper limit, $\rho = 1$, reached in (4.48) only if slacks are zero in all inputs and outputs.

4.4.2 Interpretation of SBM as a Product of Input and Output Inefficiencies

The formula for ρ in (4.48) can be transformed into

$$\rho = \left(\frac{1}{m} \sum_{i=1}^m \frac{x_{io} - s_i^-}{x_{io}} \right) \left(\frac{1}{s} \sum_{r=1}^s \frac{y_{ro} + s_r^+}{y_{ro}} \right)^{-1}.$$

The ratio $(x_{io} - s_i^-)/x_{io}$ evaluates the relative reduction rate of input i and, therefore, the first term corresponds to the mean proportional reduction rate of inputs or *input mix inefficiencies*. Similarly, in the second term, the ratio $(y_{ro} + s_r^+)/y_{ro}$ evaluates the relative proportional expansion rate of output r and $(1/s) \sum (y_{ro} + s_r^+)/y_{ro}$ is the mean proportional rate of output expansion. Its inverse, the second term, measures *output mix inefficiency*. Thus, SBM ρ can be interpreted as the ratio of mean input and output mix inefficiencies. Further, we have the theorem:

Theorem 4.7 *If DMU A dominates DMU B so that $x_A \leq x_B$ and $y_A \geq y_B$, then $\rho_A^* \geq \rho_B^*$.*

4.4.3 Solving SBM

(SBM) as formulated in (4.48) can be transformed into the program below by introducing a positive scalar variable t . See Problem 3.1 at the end of the preceding chapter.

$$(SBMt) \quad \begin{aligned} & \min_{t, \lambda, s^-, s^+} \tau = t - \frac{1}{m} \sum_{i=1}^m ts_i^- / x_{io} \\ \text{subject to} \quad & 1 = t + \frac{1}{s} \sum_{r=1}^s ts_r^+ / y_{ro} \\ & \mathbf{x}_o = X\lambda + \mathbf{s}^- \\ & \mathbf{y}_o = Y\lambda - \mathbf{s}^+ \\ & \lambda \geq 0, \mathbf{s}^- \geq 0, \mathbf{s}^+ \geq 0, t > 0. \end{aligned} \quad (4.50)$$

Now let us define

$$\mathbf{S}^- = ts^-, \mathbf{S}^+ = ts^+, \text{ and } \boldsymbol{\Lambda} = t\boldsymbol{\lambda}.$$

Then (SBMt) becomes the following linear program in t , \mathbf{S}^- , \mathbf{S}^+ , and $\boldsymbol{\Lambda}$:

$$(LP) \quad \begin{aligned} & \min \tau = t - \frac{1}{m} \sum_{i=1}^m S_i^- / x_{io} \\ \text{subject to} \quad & 1 = t + \frac{1}{s} \sum_{r=1}^s S_r^+ / y_{ro} \\ & t\mathbf{x}_o = X\boldsymbol{\Lambda} + \mathbf{S}^- \\ & t\mathbf{y}_o = Y\boldsymbol{\Lambda} - \mathbf{S}^+ \\ & \boldsymbol{\Lambda} \geq 0, \mathbf{S}^- \geq 0, \mathbf{S}^+ \geq 0, t > 0. \end{aligned} \quad (4.51)$$

Note that the choice $t > 0$ means that the transformation is reversible. Thus let an optimal solution of (LP) be

$$(\tau^*, t^*, \boldsymbol{\Lambda}^*, \mathbf{S}^{-*}, \mathbf{S}^{+*}).$$

We then have an optimal solution of (SBM) defined by,

$$\rho^* = \tau^*, \quad \boldsymbol{\lambda}^* = \boldsymbol{\Lambda}^*/t^*, \quad \mathbf{s}^{-*} = \mathbf{S}^{-*}/t^*, \quad \mathbf{s}^{+*} = \mathbf{S}^{+*}/t^*. \quad (4.52)$$

From this optimal solution, we can decide whether a DMU is *SBM-efficient* as follows:

Definition 4.6 (SBM-efficient) A DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is SBM-efficient if and only if $\rho^* = 1$.

This condition is equivalent to $\mathbf{s}^{-*} = \mathbf{0}$ and $\mathbf{s}^{+*} = \mathbf{0}$, i.e., no input excess and no output shortfall in an optimal solution.

For an SBM-inefficient DMU $(\mathbf{x}_o, \mathbf{y}_o)$, we have the expression:

$$\mathbf{x}_o = X\boldsymbol{\lambda}^* + \mathbf{s}^{-*}$$

$$\mathbf{y}_o = Y\boldsymbol{\lambda}^* - \mathbf{s}^{+*}.$$

The DMU $(\mathbf{x}_o, \mathbf{y}_o)$ can be improved and becomes efficient by deleting the input excesses and augmenting the output shortfalls. This is accomplished by the following formulae (SBM-projection):

$$\hat{\mathbf{x}}_o \Leftarrow \mathbf{x}_o - \mathbf{s}^{-*} \quad (4.53)$$

$$\hat{\mathbf{y}}_o \Leftarrow \mathbf{y}_o + \mathbf{s}^{+*}, \quad (4.54)$$

which are the same as for the Additive model. See (4.44) and (4.45).

Based on $\boldsymbol{\lambda}^*$, we define the reference set for $(\mathbf{x}_o, \mathbf{y}_o)$ as:

Definition 4.7 (Reference set) The set of indices corresponding to positive λ_j^* s is called the reference set for $(\mathbf{x}_o, \mathbf{y}_o)$.

For multiple optimal solutions, the reference set is not unique. We can, however, choose any one for our purposes.

Let R_o be the reference set designated by

$$R_o = \{ j \mid \lambda_j^* > 0 \} \quad (j \in \{1, \dots, n\}). \quad (4.55)$$

Then using R_o , we can also express $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ by,

$$\hat{\mathbf{x}}_o = \sum_{j \in R_o} \mathbf{x}_j \lambda_j^* \quad (4.56)$$

$$\hat{\mathbf{y}}_o = \sum_{j \in R_o} \mathbf{y}_j \lambda_j^*. \quad (4.57)$$

This means that $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$, a point on the efficient frontier, is expressed as a positive combination of the members of the reference set, R_o , each member of which is also efficient. See Definition 4.2.

4.4.4 SBM and the CCR Measure

In its weak efficiency form, the CCR model can be formulated as follows:

$$(CCR) \quad \min_{\theta, \mu, t^-, t^+} \theta \quad (4.58)$$

$$\text{subject to} \quad \theta x_o = X\mu + t^- \quad (4.58)$$

$$y_o = Y\mu - t^+ \quad (4.59)$$

$$\mu \geq 0, t^- \geq 0, t^+ \geq 0.$$

Now let an optimal solution of (CCR) be $(\theta^*, \mu^*, t^{*-}, t^{+*})$. From (4.58), we can derive

$$x_o = X\mu^* + t^{*-} + (1 - \theta^*)x_o \quad (4.60)$$

$$y_o = Y\mu^* - t^{+*}. \quad (4.61)$$

Let us define

$$\lambda = \mu^* \quad (4.62)$$

$$s^- = t^{*-} + (1 - \theta^*)x_o \quad (4.63)$$

$$s^+ = t^{+*}. \quad (4.64)$$

Then, (λ, s^-, s^+) is feasible for (SBM) and, by inserting the definitions in (4.63) and (4.64), its objective value can be expressed as:

$$\rho = \frac{1 - \frac{1}{m} \{ \sum_{i=1}^m t_i^{*-} / x_{io} + m(1 - \theta^*) \}}{1 + \frac{1}{s} \sum_{r=1}^s t_r^{+*} / y_{ro}} = \frac{\theta^* - \frac{1}{m} \sum_{i=1}^m t_i^{*-} / x_{io}}{1 + \frac{1}{s} \sum_{r=1}^s t_r^{+*} / y_{ro}}. \quad (4.65)$$

Evidently, the last term is not greater than θ^* . Thus, we have:

Theorem 4.8 (Tone (1997)) *The optimal SBM ρ^* is not greater than the optimal CCR θ^* .*

This theorem reflects the fact that SBM accounts for *all* inefficiencies whereas θ^* accounts only for “purely technical” inefficiencies. Notice that the coefficient $1/(m x_{io})$ of the input excesses s_i^- in ρ plays a crucial role in validating Theorem 4.8.

Conversely, for an optimal solution $(\rho^*, \lambda^*, s^{*-}, s^{+*})$ to SBM, let us transform the constraints from (4.48) into

$$\theta x_o = X\lambda^* + (\theta - 1)x_o + s^{*-} \quad (4.66)$$

$$y_o = Y\lambda^* - s^{+*}. \quad (4.67)$$

Further, we add the constraint

$$(\theta - 1)x_o + s^{*-} \geq 0. \quad (4.68)$$

Then, $(\theta, \lambda^*, t^- = (\theta - 1)x_o + s^{*-}, t^+ = s^{+*})$ is feasible for (CCR).

The relationship between CCR-efficiency and SBM-efficiency is given in the following theorem:

Theorem 4.9 (Tone (1997)) *A DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is CCR-efficient if and only if it is SBM-efficient.*

Proof. Suppose that $(\mathbf{x}_o, \mathbf{y}_o)$ is CCR-inefficient. Then, we have either $\theta^* < 1$ or $(\theta^* = 1 \text{ and } (\mathbf{t}^{-*}, \mathbf{t}^{**}) \neq (\mathbf{0}, \mathbf{0}))$. From (4.65), in both cases, we have $\rho < 1$ for a feasible solution of (SBM). Therefore, $(\mathbf{x}_o, \mathbf{y}_o)$ is SBM-inefficient.

On the other hand, suppose that $(\mathbf{x}_o, \mathbf{y}_o)$ is SBM-inefficient. Then, from Definition 4.6, $(\mathbf{s}^{-*}, \mathbf{s}^{**}) \neq (\mathbf{0}, \mathbf{0})$. By (4.66) and (4.67), $(\theta, \lambda^*, \mathbf{t}^- = (\theta - 1)\mathbf{x}_o + \mathbf{s}^{-*}, \mathbf{t}^+ = \mathbf{s}^{**})$ is feasible for (CCR), provided $(\theta - 1)\mathbf{x}_o + \mathbf{s}^{-*} \geq \mathbf{0}$. There are two cases:

(Case 1) $\theta = 1$ and $(\mathbf{t}^- = \mathbf{s}^{-*}, \mathbf{t}^+ = \mathbf{s}^{**}) \neq (\mathbf{0}, \mathbf{0})$. In this case, an optimal solution for (CCR) is CCR-inefficient.

(Case 2) $\theta < 1$. In this case, $(\mathbf{x}_o, \mathbf{y}_o)$ is CCR-inefficient.

Therefore, CCR-inefficiency is equivalent to SBM-inefficiency. Since the definitions of *efficient* and *inefficient* are mutually exclusive and collectively exhaustive, we have proved the theorem. \square

4.4.5 The Dual Program of the SBM Model

The dual program of the problem (LP) in (4.50) can be expressed as follows, with the dual variables $\xi \in R$, $\mathbf{v} \in R^m$ and $\mathbf{u} \in R^s$:

$$(DP) \quad \max_{\xi, \mathbf{v}, \mathbf{u}} \xi \quad (4.69)$$

$$\text{subject to} \quad \xi + \mathbf{v}\mathbf{x}_o - \mathbf{u}\mathbf{y}_o = 1 \quad (4.70)$$

$$-\mathbf{v}\mathbf{X} + \mathbf{u}\mathbf{Y} \leq \mathbf{0} \quad (4.71)$$

$$\mathbf{v} \geq \frac{1}{m}[1/\mathbf{x}_o] \quad (4.72)$$

$$\mathbf{u} \geq \frac{\xi}{s}[1/\mathbf{y}_o], \quad (4.73)$$

Where the notation $[1/\mathbf{x}_o]$ designates the row vector $(1/x_{1o}, \dots, 1/x_{mo})$.

Using (4.70) we can eliminate ξ and we have an equivalent program:

$$(DP') \quad \max \quad \mathbf{u}\mathbf{y}_o - \mathbf{v}\mathbf{x}_o \quad (4.74)$$

$$\text{subject to} \quad -\mathbf{v}\mathbf{X} + \mathbf{u}\mathbf{Y} \leq \mathbf{0} \quad (4.75)$$

$$\mathbf{v} \geq \frac{1}{m}[1/\mathbf{x}_o] \quad (4.76)$$

$$\mathbf{u} \geq \frac{1 - \mathbf{v}\mathbf{x}_o + \mathbf{u}\mathbf{y}_o}{s}[1/\mathbf{y}_o]. \quad (4.77)$$

The dual variables $\mathbf{v} \in R^m$ and $\mathbf{u} \in R^s$ can be interpreted as the virtual costs and prices of inputs and outputs, respectively. The dual program aims to find the optimal virtual costs and prices for the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ so that the virtual profit $\mathbf{u}\mathbf{y}_j - \mathbf{v}\mathbf{x}_j$ does not exceed zero for any DMU (including $(\mathbf{x}_o, \mathbf{y}_o)$), and

maximizes the profit $\mathbf{u}y_o - \mathbf{v}x_o$ for the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ concerned. Apparently, the optimal profit is at best zero and hence $\xi = 1$ for the SBM efficient DMUs. Constraints (4.76) and (4.77) restrict the feasible \mathbf{v} and \mathbf{u} to the positive orthant.

4.4.6 Oriented SBM Models

The input (output)-oriented SBM model can be defined by neglecting the denominator (numerator) of the objective function (4.48) of (SBM). Thus, the efficiency values ρ_I^* and ρ_O^* can be obtained as follows:

[Input-oriented SBM Model]

$$(SBM - I) \quad \rho_I^* = \min_{\lambda, s^-} 1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{io} \quad (4.78)$$

subject to $\mathbf{x}_o = X\lambda + s^-$
 $\mathbf{y}_o \leq Y\lambda$
 $\lambda \geq 0, s^- \geq 0.$

[Output-oriented SBM Model]

$$(SBM - O) \quad \rho_O^* = \min_{\lambda, s^+} \frac{1}{1 + \frac{1}{s} \sum_{r=1}^s s_r^+ / y_{ro}} \quad (4.79)$$

subject to $\mathbf{x}_o \geq X\lambda$
 $\mathbf{y}_o = Y\lambda - s^+$
 $\lambda \geq 0, s^+ \geq 0.$

This leads to:

Theorem 4.10

$$\rho_I^* \geq \rho^* \text{ and } \rho_O^* \geq \rho^*. \quad (4.80)$$

The input-oriented SBM is substantially equivalent to what is called the “Russell input measure of technical efficiency.”⁹

4.4.7 A Weighted SBM Model

We can assign weights to inputs and outputs corresponding to the relative importance of items as follows:

$$\rho = \frac{1 - \frac{1}{m} \sum_{i=1}^m w_i^- s_i^- / x_{io}}{1 + \frac{1}{s} \sum_{r=1}^s w_r^+ s_r^+ / y_{ro}}, \quad (4.81)$$

with

$$\sum_{i=1}^m w_i^- = m \text{ and } \sum_{r=1}^s w_r^+ = s. \quad (4.82)$$

The weights should reflect the intentions of the decision-maker. If all outputs are in the same unit, e.g., dollars, a conventional scheme is that:

$$w_r^+ = s \sum_{j=1}^n y_{rj} / \sum_{j=1}^n \sum_{k=1}^s y_{kj}. \quad (4.83)$$

This weight selection reflects the importance of the output r being proportional to its average magnitude. The input weights can be determined analogously.

4.4.8 Decomposition of Inefficiency

Using optimal slacks (s^{-*}, s^{+*}) for (SBM) in (4.48), we can decompose the SBM score ρ^* as follows:

$$\rho^* = \frac{1 - \sum_{i=1}^m \alpha_i}{1 + \sum_{r=1}^s \beta_r} \quad (4.84)$$

$$\text{where } \alpha_i = \frac{1}{m} \frac{s_i^{-*}}{x_{io}} \quad (i = 1, \dots, m) \quad (4.85)$$

$$\beta_r = \frac{1}{s} \frac{s_r^{+*}}{y_{ro}}. \quad (r = 1, \dots, s) \quad (4.86)$$

This decomposition is informative for evaluating the sources and magnitudes of inefficiency regarding the respective inputs and outputs for the DMU_o . (See Problem 4.12 for an example of this decomposition.)

4.4.9 Numerical Example (SBM)

We illustrate SBM using an example. Table 4.3 exhibits data for eight DMUs using two inputs (x_1, x_2) to produce a single output ($y = 1$), along with CCR, SBM scores, slacks and reference set. Although DMUs F and G have full CCR-score ($\theta^* = 1$), they have slacks compared to C and this is reflected by drops in the SBM scores to $\rho_F^* = 0.9$ and $\rho_G^* = 0.83333$. Also, the SBM scores of inefficient DMUs A , B and H dropped slightly from the CCR scores due to their slacks. Thus, the SBM measure reflects not only the weak efficiency values in θ^* but also the other (slack) inefficiencies as well.

4.5 A HYBRID MEASURE OF EFFICIENCY (HYBRID)

There are two types of measures or approaches in DEA: *radial* and *non-radial*. Differences exist in the characterization of input or output items. Suppose that there are four inputs x_1, x_2, x_3 and x_4 , in the problem of interest, where x_1 and x_2 are radial, and x_3 and x_4 are non-radial. That is, (x_1, x_2) are subject to change proportionally, such as $(\alpha x_1, \alpha x_2)$ (with $\alpha > 0$), while x_3 and x_4 are subject to change non-radially. These differences should be reflected in the evaluation of efficiency. The radial input part (x_1, x_2) satisfies the efficiency status if there is no proportionally reduced input $(\alpha x_1, \alpha x_2)$ (with $\alpha < 1$) that can produce at least the observed outputs. The non-radial input part x_3 (x_4)

Table 4.3. Data and Results of CCR and SBM

DMU	Data			θ^*	CCR				SBM				Mix Eff. ^a
	x_1	x_2	y		$\rho^* = \rho_I^*$	Ref.	s_1^{-*}	s_2^{-*}	s^{+*}				
A	4	3	1	0.857	0.833	D	0	1	0				0.972
B	7	3	1	0.632	0.619	D	3	1	0				0.98
C	8	1	1	1	1	C	0	0	0				1
D	4	2	1	1	1	D	0	0	0				1
E	2	4	1	1	1	E	0	0	0				1
F	10	1	1	1	0.9	C	2	0	0				0.9
G	12	1	1	1	0.833	C	4	0	0				0.833
H	10	1.5	1	0.75	0.733	C	2	0.5	0				0.978

^a: Mix Eff. = ρ^*/θ^* . See Section 5.8.2 for detail.

satisfies the efficiency status if there is no reduced x_3 (x_4) that can produce the observed outputs. Analogously, the output part can be divided into the radial and non-radial outputs.

The radial approach is represented by the CCR and BCC models. Its shortcoming is that it neglects the non-radial input/output slacks. The non-radial approach SBM deals with slacks directly, but it neglects the radial characteristics of inputs and/or outputs.

In this section, we integrate these approaches in a unified framework and present a hybrid measure of efficiency (Hybrid). ¹⁰

4.5.1 A Hybrid Measure

Let the observed input and output data matrices be $X \in R_+^{m \times n}$ and $Y \in R_+^{s \times n}$, respectively, where n , m and s designate the number of DMUs, inputs and outputs. We decompose the input matrix into the radial part, $X^R \in R_+^{m_1 \times n}$, and non-radial part, $X^{NR} \in R_+^{m_2 \times n}$, with $m = m_1 + m_2$, as follows:

$$X = \begin{pmatrix} X^R \\ X^{NR} \end{pmatrix}. \quad (4.87)$$

Analogously, we decompose the output matrix Y into the radial part, $Y^R \in R_+^{s_1 \times n}$ and the non-radial part, $Y^{NR} \in R_+^{s_2 \times n}$, with $s = s_1 + s_2$, as follows:

$$Y = \begin{pmatrix} Y^R \\ Y^{NR} \end{pmatrix}. \quad (4.88)$$

We assume that the data set is positive, i.e., $X > 0$ and $Y > 0$. The production possibility set P is defined by

$$P = \{(x, y) | x \geq X\lambda, y \leq Y\lambda, \lambda \geq 0\}, \quad (4.89)$$

where λ is a nonnegative vector in R^n . (We can impose some constraints on λ , such as $\sum_{j=1}^n \lambda_j = 1$, which is the variable returns to scale model.)

We consider an expression for describing a certain DMU $(x_o, y_o) = (x_o^R, x_o^{NR}, y_o^R, y_o^{NR}) \in P$ as

$$\begin{aligned}\theta x_o^R &= X^R \lambda + s^{R-} \\ x_o^{NR} &= X^{NR} \lambda + s^{NR-} \\ \phi y_o^R &= Y^R \lambda - s^{R+} \\ y_o^{NR} &= Y^{NR} \lambda - s^{NR+},\end{aligned}\tag{4.90}$$

with $\theta \leq 1$, $\phi \geq 1$, $\lambda \geq 0$, $s^{R-} \geq 0$, $s^{NR-} \geq 0$, $s^{R+} \geq 0$, $s^{NR+} \geq 0$. The vectors $s^{R-} \in R^{m_1}$ and $s^{NR-} \in R^{m_2}$ indicate the *excesses* for the radial and non-radial inputs, respectively, while $s^{R+} \in R^{s_1}$ and $s^{NR+} \in R^{s_2}$ indicate the *shortfalls* for the radial and non-radial outputs, respectively. They are called *slack*s.

As such, $\theta = 1$, $\phi = 1$, $\lambda_o = 1$, $\lambda_j = 0 (\forall j \neq o)$, with all slacks being zero is a feasible expression. Based on the expression (4.90), we define an index ρ as follows:

$$\rho = \frac{1 - \frac{m_1}{m}(1 - \theta) - \frac{1}{m} \sum_{i=1}^{m_2} s_i^{NR-} / x_{io}^{NR}}{1 + \frac{s_1}{s}(\phi - 1) + \frac{1}{s} \sum_{r=1}^{s_2} s_r^{NR+} / y_{ro}^{NR}}.\tag{4.91}$$

This index ρ is designed so that it is decreasing with respect to decreases in θ and increases in ϕ , $s_i^{NR-} (\forall i)$ and $s_r^{NR+} (\forall r)$, but is not affected by s^{R-} and s^{R+} directly, reflecting free disposability of these radial slacks. This index is also units invariant, i.e., invariant with respect to the measurement units of the data.

The hybrid efficiency status of the DMU $(x_o, y_o) = (x_o^R, x_o^{NR}, y_o^R, y_o^{NR}) \in P$ is defined as follows:

Definition 4.8 (Hybrid efficient status) *The DMU (x_o, y_o) is hybrid efficient if and only if $\rho = 1$ for every feasible expression of (4.90), i.e., $\theta = 1$, $\phi = 1$, $s^{NR-} = 0$, $s^{NR+} = 0$.*

This status can be ascertained by solving the following program with the variables θ , ϕ , λ , s^{NR-} , s^{NR+} .

$$\begin{aligned}(\text{Hybrid}) \quad \rho^* &= \min \frac{1 - \frac{m_1}{m}(1 - \theta) - \frac{1}{m} \sum_{i=1}^{m_2} s_i^{NR-} / x_{io}^{NR}}{1 + \frac{s_1}{s}(\phi - 1) + \frac{1}{s} \sum_{r=1}^{s_2} s_r^{NR+} / y_{ro}^{NR}} \\ \text{subject to} \quad &\theta x_o^R \geq X^R \lambda \\ &x_o^{NR} = X^{NR} \lambda + s^{NR-} \\ &\phi y_o^R \leq Y^R \lambda \\ &y_o^{NR} = Y^{NR} \lambda - s^{NR+} \\ &\theta \leq 1, \phi \geq 1, \lambda \geq 0, s^{NR-} \geq 0, s^{NR+} \geq 0.\end{aligned}\tag{4.92}$$

Let an optimal solution for this program be $(\theta^*, \phi^*, \lambda^*, s^{NR-*}, s^{NR+*})$. Then we have the following theorem:

Theorem 4.11 *The DMU (x_o, y_o) is hybrid-efficient if and only if $\rho^* = 1$, i.e., $\theta^* = 1, \phi^* = 1, s^{NR-*} = \mathbf{0}, s^{NR+*} = \mathbf{0}$.*

The (*Hybrid*) can be transformed into a linear program using the Charnes-Cooper transformation (Charnes and Cooper (1962)) as follows:

$$(LP) \quad \tau^* = \min t - \frac{m_1}{m}(t - \Theta) - \frac{1}{m} \sum_{i=1}^{m_2} S_i^{NR-} / x_{io}^{NR} \quad (4.93)$$

$$\text{subject to } t + \frac{s_1}{s}(\Phi - t) + \frac{1}{s} \sum_{r=1}^{s_2} S_r^{NR+} / y_{ro}^{NR} = 1$$

$$\Theta x_o^R \geq X^R \Lambda$$

$$tx_o^{NR} = X^{NR} \Lambda + S^{NR-}$$

$$\Phi y_o^R \leq Y^R \Lambda$$

$$ty_o^{NR} = Y^{NR} \Lambda - S^{NR+},$$

$$\Theta \leq t, \Phi \geq t, \Lambda \geq 0, S^{NR-} \geq 0, S^{NR+} \geq 0.$$

Let an optimal solution of (LP) be $(t^*, \Theta^*, \Phi^*, \Lambda^*, S^{NR-*}, S^{NR+*})$. Then we have an optimal solution of (*Hybrid*) as defined by:

$$\begin{aligned} \rho^* &= \tau^*, \theta^* = \Theta^*/t^*, \phi^* = \Phi^*/t^*, \\ \lambda^* &= \Lambda^*/t^*, s^{NR-*} = S^{NR-*}/t^*, s^{NR+*} = S^{NR+*}/t^*. \end{aligned} \quad (4.94)$$

For a hybrid-inefficient DMU, i.e., $\rho^* < 1$, the hybrid-projection is given by:

$$\bar{x}_o^R \leftarrow \theta^* x_o^R \quad (4.95)$$

$$\bar{x}_o^{NR} \leftarrow x_o^{NR} - s^{NR-*} \quad (4.96)$$

$$\bar{y}_o^R \leftarrow \phi^* y_o^R \quad (4.97)$$

$$\bar{y}_o^{NR} \leftarrow y_o^{NR} + s^{NR+*}. \quad (4.98)$$

We notice that the radial slacks s^{R-*} and s^{R+*} , if they exist, are not accounted for in the above projection, since they are assumed to be freely disposable and have no effect on efficiency evaluation.

Theorem 4.12 (Tone (2004)) *The projected DMU $(\bar{x}_o^R, \bar{x}_o^{NR}, \bar{y}_o^R, \bar{y}_o^{NR})$ is hybrid-efficient.*

See Problem 4.6 for a proof.

4.5.2 Decomposition of Inefficiency

Using the optimal solution $(\theta^*, \phi^*, s^{NR-*}, s^{NR+*})$, we can decompose the hybrid efficiency indicator ρ^* into four factors as follows:

$$\text{Radial input inefficiency: } \alpha_1 = \frac{m_1}{m}(1 - \theta^*) \quad (4.99)$$

$$\text{Non-radial input inefficiency: } \alpha_2 = \frac{1}{m} \sum_{i=1}^{m_2} s_i^{NR-*} / x_{io}^{NR} \quad (4.100)$$

$$\text{Radial output inefficiency: } \beta_1 = \frac{s_1}{s} (\phi^* - 1) \quad (4.101)$$

$$\text{Non-radial output inefficiency: } \beta_2 = \frac{1}{s} \sum_{r=1}^{s_2} s_r^{NR+*} / y_{ro}^{NR}. \quad (4.102)$$

We also define input and output inefficiencies as:

$$\text{Input inefficiency: } \alpha = \alpha_1 + \alpha_2 \quad (4.103)$$

$$\text{Output inefficiency: } \beta = \beta_1 + \beta_2. \quad (4.104)$$

Thus, ρ^* can be expressed as:

$$\rho^* = \frac{1 - \alpha}{1 + \beta} = \frac{1 - \alpha_1 - \alpha_2}{1 + \beta_1 + \beta_2}. \quad (4.105)$$

This expression is useful for finding the sources of inefficiency and the magnitude of their influence on the efficiency score ρ^* .

4.5.3 Comparisons with the CCR and SBM Models

We can define the CCR and SBM models as special cases of the Hybrid model and compare their efficiency values.

The Hybrid model can be transformed into the CCR or SBM model by setting all inputs and outputs as *radial* or *non-radial*, respectively. Thus, we can define:

$$(CCR) \quad \rho_{CCR}^* = \min \frac{\theta}{\phi} \quad (4.106)$$

$$\text{subject to } \theta \mathbf{x}_o \geq X \boldsymbol{\lambda}$$

$$\phi \mathbf{y}_o \leq Y \boldsymbol{\lambda}$$

$$\theta \leq 1, \phi \geq 1, \boldsymbol{\lambda} \geq \mathbf{0}.$$

$$(SBM) \quad \rho_{SBM}^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{io}}{1 + \frac{1}{s} \sum_{r=1}^s s_r^+ / y_{ro}} \quad (4.107)$$

$$\text{subject to } \mathbf{x}_o = X \boldsymbol{\lambda} + \mathbf{s}^-$$

$$\mathbf{y}_o = Y \boldsymbol{\lambda} - \mathbf{s}^+$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}.$$

In a similar vein, as described in Section 4.4.6, we can define the oriented CCR or SBM models, i.e., CCR-I, CCR-O, or Hybrid-I, Hybrid-O, by neglecting the output or input side efficiencies. Between the CCR, CCR-I and CCR-O models we have:

Lemma 4.2

$$\rho_{CCR}^* = \rho_{CCR-I}^* = \rho_{CCR-O}^*. \quad (4.108)$$

See Problem 4.9 for a proof.

However, this lemma is no longer true when we deal with the BCC model, i.e., the radial and VRS model. Usually, ρ_{BCC}^* , ρ_{BCC-I}^* , and ρ_{BCC-O}^* are different each other.

Among these indices, we have the following relationship.

Theorem 4.13 (Tone (2004))

$$\begin{aligned}\rho_{CCR}^* &\geq \rho^* \geq \rho_{SBM}^* \\ \rho_{CCR-I}^* &\geq \rho_I^* \geq \rho_{SBM-I}^* \\ \rho_{CCR-O}^* &\geq \rho_O^* \geq \rho_{SBM-O}^* \\ \rho_{SBM-I}^* &\geq \rho_{SBM}^*, \quad \rho_{SBM-O}^* \geq \rho_{SBM}^* \\ \rho_{BCC-I}^* &\geq \rho_{BCC}^*, \quad \rho_{BCC-O}^* \geq \rho_{BCC}^*,\end{aligned}\quad (4.109)$$

where ρ^* , ρ_I^* and ρ_O^* represent the optimal scores of (Hybrid), (Hybrid-I) and (Hybrid-O), respectively.

See Problem 4.10 for a proof.

4.5.4 An Illustrative Example

Table 4.4 exhibits a simple example consisting of six DMUs, each having two radial inputs (x_1 , x_2), a single non-radial input (x_3), two radial outputs (y_1 , y_2), and a single non-radial output (y_3). We measured the efficiency score using the Hybrid, CCR and SBM models and obtained the results as shown in the last three columns of Table 4.4, where CCR and SBM assume all variables are Radial and Non-radial, respectively.

Table 4.4. A Comparison: Hybrid, CCR and SBM

DMU	Input			Output			Efficiency score		
	Radial	Non-radial		Radial	Non-radial		Hybrid	CCR	SBM
	x_1	x_2	x_3	y_1	y_2	y_3			
A	1	1	1	2	2	2	1	1	1
B	1	2	1	2	2	2	1	1	0.833
C	1	1	2	1	2	2	0.833	1	0.625
D	1	1	1	1	2	1	0.75	1	0.6
E	1	1	1	2	1	0.5	0.5	1	0.429
F	2	2	2	0.5	0.5	1	0.15	0.25	0.15

Observing the data set, we can see that the DMU A is the only superstar in the sense that other DMUs have input excesses or output shortfalls against

A. The non-radial SBM captures these drawbacks effectively and gives the full score unity only to A, whereas others are judged to be inefficient. On the other hand, the radial CCR model neglects these slacks and assigns the full efficiency score unity to all DMUs except F. For example, DMU B has 1 unit excess in x_2 against A. However, B has $x_1 = x_3 = 1$ and $y_1 = y_2 = y_3 = 1$ which are the same as the superstar A. Hence, this excess is neglected and B has the CCR-efficient status. DMUs C, D and E are CCR-efficient by the same reasoning. This is a remarkable characteristics of the *radial* models.

The Hybrid model gives the score between CCR and SBM. It accounts for the *non-radial* excesses and shortfalls.

Table 4.5 exhibits the decomposition of the efficiency score by means of the expression (4.105). From this table, we can identify the sources of inefficiency and their magnitudes. For example, DMU C's inefficiency is caused by the input inefficiency ($\alpha = 0.167$), which stems from the non-radial input inefficiency ($\alpha_2 = 0.167$). DMU F's inefficiency can be attributed to the output inefficiency ($\beta = 5.67$), which stems from the radial output inefficiency ($\beta_1 = 4.67$) and the non-radial output inefficiency ($\beta_2 = 1$). This tallies with the value $\phi_F = 8$ in the table. $\theta = 1$ and $\phi = 1$ indicate the absence of radial inefficiency in inputs and outputs, respectively.

Table 4.5. Measures of Inefficiency: (*Hybrid*)

DMU	Score	α	Input		β	Output		Input θ	Output ϕ
			α_1	α_2		β_1	β_2		
A	1	0	0	0	0	0	0	1	1
B	1	0	0	0	0	0	0	1	1
C	0.833	0.167	0	0.167	0	0	0	1	1
D	0.75	0	0	0	0.333	0	0.333	1	1
E	0.5	0	0	0	1	0	1	1	1
F	0.15	0	0	0	5.67	4.67	1	1	8

4.6 RUSSELL MEASURE MODELS

We now introduce a model due to R.R. Russell. Described as the “Russell Measure Model” by Färe and Lovell(1978),¹¹ their formulation is difficult to compute. We therefore turn to a more recent development due to Pastor, Ruiz and Sirvent.¹² This model is

$$R(\mathbf{x}_o, \mathbf{y}_o) = \min_{\theta, \phi, \lambda} \frac{\sum_{i=1}^m \theta_i/m}{\sum_{r=1}^s \phi_r/s} \quad (4.110)$$

subject to

$$\begin{aligned}\theta_i x_{io} &\geq \sum_{j=1}^n x_{ij} \lambda_j, \quad i = 1, \dots, m \\ \phi_r y_{ro} &\leq \sum_{j=1}^n y_{rj} \lambda_j, \quad r = 1, \dots, s \\ 0 &\leq \lambda_j \quad \forall j \\ 0 &\leq \theta_i \leq 1; \quad 1 \leq \phi_r \quad \forall i, r.\end{aligned}$$

Pastor *et al.* refer to this as the “Enhanced Russell Graph Measure of Efficiency” but we shall refer to it as ERM (Enhanced Russell Measure). See Färe, Grosskopf and Lovell (1985)¹³ for the meaning of “graph measure.” Such measures are said to be “closed” so $R(\mathbf{x}_o, \mathbf{y}_o)$ includes all inefficiencies that the model can identify. In this way we avoid limitations of the radial measures which cover only some of the input or output inefficiencies and hence measure only “weak efficiency.”

The “inclusive” (=closure) property is shared by SBM. In fact SBM and ERM are related as in the following theorem,

Theorem 4.14 : *ERM as formulated in (4.110) and SBM as formulated in (4.48) are equivalent in that λ_j^* values that are optimal for one are also optimal for the other.*

Proof : As inspection makes clear, a necessary condition for optimality of ERM is that the constraints in (4.110) must be satisfied as equalities. Hence we can replace those constraints with

$$\begin{aligned}\theta_i &= \sum_{j=1}^n x_{ij} \lambda_j / x_{io}, \quad i = 1, \dots, m \\ \phi_r &= \sum_{j=1}^n y_{rj} \lambda_j / y_{ro}, \quad r = 1, \dots, s.\end{aligned}\tag{4.111}$$

Following Pastor *et al.* or Bardhan *et al.*,¹⁴ we next set

$$\begin{aligned}\theta_i &= \frac{x_{io} - s_i^-}{x_{io}} = 1 - \frac{s_i^-}{x_{io}}, \quad i = 1, \dots, m \\ \phi_r &= \frac{y_{ro} + s_r^+}{y_{ro}} = 1 + \frac{s_r^+}{y_{ro}}, \quad r = 1, \dots, s.\end{aligned}\tag{4.112}$$

Substituting these values in (4.111) produces

$$\begin{aligned}1 &= \frac{\sum_{j=1}^n x_{ij} \lambda_j}{x_{io}} + \frac{s_i^-}{x_{io}}, \quad i = 1, \dots, m \\ 1 &= \frac{\sum_{j=1}^n y_{rj} \lambda_j}{y_{ro}} + \frac{s_r^+}{y_{ro}}, \quad r = 1, \dots, s\end{aligned}\tag{4.113}$$

or

$$\begin{aligned} x_{io} &= \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s \end{aligned} \quad (4.114)$$

which are the same as the constraints for SBM in (4.48).

Turning to the additional conditions, $0 \leq \theta_i \leq 1$ and $1 \leq \phi_r$ we again substitute from (4.112) to obtain $0 \leq s_i^- \leq x_{io}$ and $0 \leq s_r^+$. The condition $s_i^- \leq x_{io}$ is redundant since it is satisfied by the first set of inequalities in (4.110). Hence, we only have non-negativity for all slacks, the same as in (4.48).

Now turning to the objective in (4.110) we once more substitute from (4.112) to obtain

$$\frac{s}{m} \frac{\sum_{i=1}^m \theta_i}{\sum_{r=1}^s \phi_r} = \frac{s}{m} \frac{\sum_{i=1}^m \left(1 - \frac{s_i^-}{x_{io}}\right)}{\sum_{r=1}^s \left(1 + \frac{s_r^+}{y_{ro}}\right)} = \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}}} \quad (4.115)$$

which is the same as the objective for SBM. Using the relations in (4.112) we may therefore use SBM to solve ERM or *vice versa*. \square

4.7 SUMMARY OF THE BASIC DEA MODELS

In Table 4.6, we summarized some important topics for consideration in choosing between basic DEA models. In this table, ‘Semi-p’ (=semipositive) means nonnegative with at least one positive element in the data for each DMU, and ‘Free’ permits negative, zero or positive data. Although we have developed some DEA models under the assumption of positive data set, this assumption can be relaxed as exhibited in the table. For example, in the BCC-I (-O) model, outputs (inputs) are free due to the translation invariance theorem. In the case of SBM, nonpositive outputs can be replaced by a very small positive number and nonpositive input terms can be neglected for consideration in the objective function. The θ^* of the output oriented models (CCR-O and BCC-O) is the reciprocal of $\eta^*(\geq 1)$. ‘Tech. or Mix’ indicates whether the model measures ‘technical efficiency’ or ‘mix efficiency’. ‘CRS’ and ‘VRS’ mean *constant* and *variable* returns to scale, respectively. The returns to scale of ADD and SBM depends on the added convexity constraint $e\lambda = 1$.

Model selection is one of the problems to be considered in DEA up to, and including, choices of multiple models to test whether or not a result is dependent on the models (or methods) used.¹⁵ Although other models will be developed in succeeding chapters, we will mention here some of the considerations to be taken into account regarding model selection.

1. The Shape of the Production Possibility Set.

The CCR model is based on the assumption that *constant* returns to scale

Table 4.6. Summary of Model Characteristics

Model		CCR-I	CCR-O	BCC-I	BCC-O	ADD	SBM(Hybrid)
Data	X	Semi-p	Semi-p	Semi-p	Free	Free	Semi-p
	Y	Free	Free	Free	Semi-p	Free	Free
Trans.	X	No	No	No	Yes	Yes ^a	No
	Y	No	No	Yes	No	Yes ^a	No
Invariance		Yes	Yes	Yes	Yes	No	Yes
θ^*		[0, 1]	[0, 1]	(0, 1]	(0, 1]	No	[0, 1]
Tech. or Mix		Tech.	Tech.	Tech.	Tech.	Mix	Mix
Returns to Scale		CRS	CRS	VRS	VRS	C(V)RS ^b	C(V)RS

^a: The Additive model is translation invariant only when the convexity constraint is added.

^b: C(V)RS means Constant or Variable returns to scale according to whether or not the convexity constraint is included.

prevails at the efficient frontiers, whereas the BCC and Additive models assume variable returns to scale frontiers, i.e., *increasing*, *constant* and *decreasing* returns to scale.¹⁶ If preliminary surveys on the production functions identify a preferable choice by, say, such methods as linear regression analysis, (e.g., a Cobb-Douglas type) or expert opinions, then we can choose a DEA model that fits the situation. However, we should bear in mind that conventional regression-based methods deal with single output and multiple input cases, while DEA models analyze multiple outputs and multiple inputs correspondences.

2. Input or Output Oriented?

One of the main purposes of a DEA study is to project the inefficient DMUs onto the production frontiers, e.g., the CCR-projection and the BCC-projection, among others. There are three directions, one called *input-oriented* that aims at reducing the input amounts by as much as possible while keeping at least the present output levels, and the other, called *output-oriented*, maximizes output levels under at most the present input consumption. There is a third choice, represented by the Additive and SBM models that deal with the input excesses and output shortfalls simultaneously in a way that jointly maximizes both. If achievement of efficiency, or failure to do so, is the only topic of interest, then these different models will all yield the same result inssofar as technical and mix inefficiency is concerned. However, we need to note that the Additive and BCC models may give different estimates when inefficiencies are present. Moreover, as we shall see in the next chapter, the CCR and BCC models differ in that the former evaluates scale as well as technical inefficiencies simultaneously whereas the latter evaluates the two in a fashion that identifies them separately.

3. Translation Invariance.

As is seen from Table 4.6, we can classify the models according to whether or not they use the efficiency measure θ^* . It is to be noted that θ^* is measured in a manner that depends on the coordinate system of the data set. On the other hand, the θ^* -free models, such as the Additive model, are essentially coordinate-free and translation invariant. They evaluate the efficiency of a DMU by the l_1 -metric distance from the efficient frontiers and are invariant with respect to the translation of the coordinate system. Although they supply information on the projection of the inefficient DMUs onto the production frontiers, they lack a one-dimensional efficiency measure like θ^* . The SBM model developed in Section 4.4 is designed to overcome such deficiencies in both the CCR-type and Additive models. However, SBM is not translation invariant.

4. Number of Input and Output Items.

Generally speaking, if the number of DMUs (n) is less than the combined number of inputs and outputs ($m + s$), a large portion of the DMUs will be identified as efficient and efficiency discrimination among DMUs is questionable due to an inadequate number of degrees of freedom. See the opening discussion in Chapter 9. Hence, it is desirable that n exceed $m + s$ by several times. A rough rule of thumb in the envelopment model is to choose n (= the number of DMUs) equal to or greater than $\max\{m \times s, 3 \times (m + s)\}$. The selection of input and output items is crucial for successful application of DEA. We therefore generally recommend a process of selecting a small set of input and output items at the beginning and gradually enlarging the set to observe the effects of the added items. In addition, other methods, e.g., the assurance region method, the cone ratio model and others (that will be presented in the succeeding chapters) lead themselves to a sharper discrimination among DMUs.

5. Try Different Models.

If we cannot identify the characteristics of the production frontiers by preliminary surveys, it may be risky to rely on only one particular model. If the application has important consequences it is wise to try different models and methods, compare results and utilize expert knowledge of the problem, and possibly try other devices, too, before arriving at a definitive conclusion. See the discussion in Chapter 9 dealing with the use of statistical regression and DEA to cross check each other.

4.8 SUMMARY OF CHAPTER 4

In this chapter, we introduced several DEA models.

1. The CCR and BCC models which are radial measures of efficiency that are either input oriented or output oriented.
2. The Additive models which identify input excesses and output shortfalls simultaneously.

3. The slacks-based measure of efficiency (SBM) which uses the Additive model and provides a scalar measure ranging from 0 to 1 that encompasses all of the inefficiencies that the model can identify.
4. The hybrid measure of efficiency integrates both the radial and non-radial measures in a unified manner.
5. We investigated the problem of model choice. All remaining chapters will serve as a complement in treating this subject.
6. The translation invariance of the Additive model was introduced. This property of the Additive model (with convexity constraint) was identified by Ali and Seiford (1990) and was extended by Pastor (1996).¹⁷ See, however, Thrall (1996)¹⁸ for limitations involved in relying on this invariance property.

4.9 NOTES AND SELECTED BIBLIOGRAPHY

Translation invariance was first shown by Ali and Seiford (1990) and extended by Pastor *et al.* (1999). The SBM model was introduced by Tone (1997, 2001). The hybrid measure of efficiency was proposed by Tone (2004).

4.10 APPENDIX: FREE DISPOSAL HULL (FDH) MODELS

Another model which has received a considerable amount of research attention is the FDH (Free Disposal Hull) model as first formulated by Deprins, Simar and Tulkens (1984)¹⁹ and developed and extended by Tulkens and his associates at the University of Louvain in Belgium.²⁰ The basic motivation is to ensure that efficiency evaluations are effected from only actually observed performances. Points like Q in Figure 3.2, for example, are not allowed because they are derived and not actually observed performances. Hence they are hypothetical.

Figure 4.7 provides an example of what is referred to as the Free Disposal Hull for five DMUs using two inputs in amounts x_1 and x_2 to produce a single output in amount $y = 1$.

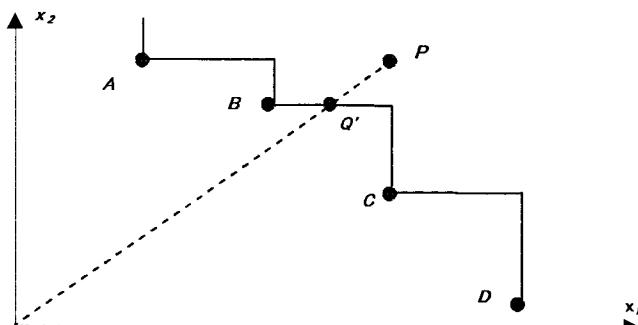


Figure 4.7. FDH Representation

The boundary of the set and its connection represents the “hull” defined as the “smallest set” that encloses all of the production possibilities that can be generated from the observations. Formally,

$$P_{FDH} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq \mathbf{x}_j, \mathbf{y} \leq \mathbf{y}_j, \mathbf{x}, \mathbf{y} \geq \mathbf{0}, j = 1, \dots, n\}$$

where $\mathbf{x}_j (\geq \mathbf{0}), \mathbf{y}_j (\geq \mathbf{0})$ are actually observed performances for $j = 1, \dots, n$ DMUs. In words, a point is a member of the production possibility set if all of its input coordinates are at least as large as their corresponds in the vector of observed values \mathbf{x}_j for any $j = 1, \dots, n$ and if their output coordinates are no greater than their corresponds in the vectors \mathbf{y}_j of observed values for this same j .

This gives rise to the staircase (or step) function which we have portrayed by the solid line in the simple two-input one-output example of Figure 4.7. No point below this solid line has the property prescribed for P_{FDH} . Moreover this boundary generates the smallest set with these properties. For instance, connecting points B and C in the manner of Figure 3.2 in Chapter 3 would generate the boundary of a bigger production possibility set. Tulkens and his associates use an algorithm that eliminates all dominated points as candidates for use in generating the FDH. This algorithm proceeds in pairwise comparison fashion as follows: Let DMU $_k$ with coordinate $\mathbf{x}_k, \mathbf{y}_k$ be a candidate. If for any DMU $_j$ we have $\mathbf{x}_j \leq \mathbf{x}_k$ or $\mathbf{y}_j \geq \mathbf{y}_k$ with either $\mathbf{x}_j \neq \mathbf{x}_k$ or $\mathbf{y}_j \neq \mathbf{y}_k$ then DMU $_k$ is dominated (strictly) and removed from candidacy. Actually this can be accomplished more simply by using the following mixed integer programming formulation,

$$\begin{aligned} \min \quad & \theta \\ \text{subject to} \quad & \theta \mathbf{x}_o - X \boldsymbol{\lambda} \geq \mathbf{0} \\ & \mathbf{y}_o - Y \boldsymbol{\lambda} \leq \mathbf{0} \\ & \mathbf{e} \boldsymbol{\lambda} = 1, \quad \lambda_j \in \{0, 1\} \end{aligned} \tag{4.116}$$

where X and Y contain the given input and output matrices and $\lambda_j \in \{0, 1\}$ means that the components of $\boldsymbol{\lambda}$ are constrained to be bivalent. That is, they must all have values of zero or unity so that together with the condition $\mathbf{e} \boldsymbol{\lambda} = 1$ one and only one of the performances actually observed can be chosen. This approach was first suggested in Bowlin *et al.*(1984)²¹ where it was coupled with an additive model to ensure that the “most dominant” of the non-dominated DMUs was chosen for making the indicated efficiency evaluation. In the case of Figure 4.7, the designation would thus have been A rather than B or C in order to maximize the sum of the slacks s_1^- and s_2^- when evaluating P . However, this raises an issue because the results may depend on the units of measure employed. One way to avoid this difficulty is to use the radial measure represented by $\min \theta = \theta^*$. This course, as elected by Tulkens and associates, would yield the point Q' shown in Figure 4.7. This, however, leaves the slack in going from Q' to B unattended, which brings into play the assumption of “Free Disposal” that is incorporated in the name “Free Disposal Hull.” As noted earlier this

means that nonzero slacks are ignored or, equivalently, weak efficiency suffices because the slacks do not appear in the objective — or, equivalently, they are present in the objective with zero coefficients assigned to them.

One way to resolve all of these issues is to return to our slacks-based measure (*SBM*) as given in (4.48) or its linear programming (*LP*) equivalent in (4.51) with the conditions $\lambda_j \in \{0, 1\}$ and $e\lambda = 1$ adjoined. This retains the advantage of the additive model which (a) selects an “actually observed performance” and (b) provides a measure that incorporates all of the thus designated inefficiencies. Finally, it provides a measure which is units invariant as well. See R.M. Thrall (1999)²² for further discussions of FDH and its limitations.

4.11 RELATED DEA-SOLVER MODELS FOR CHAPTER 4

BCC-I (The input-oriented Banker-Charnes-Cooper model).

This code solves the BCC model expressed by (4.2)-(4.6). The data format is the same as for the CCR model. The “Weight” sheet includes the optimal value of the multiplier u_0 corresponding to the constraint $\sum_{j=1}^n \lambda_j = 1$, as well as v^* and u^* . In addition to the results which are similar to the CCR case, this model finds the returns-to-scale characteristics of DMUs in the “RTS” sheet. For inefficient DMUs, we identify returns to scale with the projected point on the efficient frontier. BCC-I uses the projection formula (4.17)-(4.18).

BCC-O (The output-oriented BCC model).

This code solves the output-oriented BCC model expressed in (4.22)-(4.26). The optimal expansion rate η_B^* is displayed in “Score” by its inverse in order to facilitate comparisons with the input-oriented case. Usually, the efficiency score differs in both cases for inefficient DMUs. This model also finds the returns-to-scale characteristics of DMUs in the “RTS” sheet. For inefficient DMUs, we identify returns to scale with the projected point on the efficient frontier. BCC-O uses the BCC version of the formula (3.74)-(3.15). The returns-to-scale characteristics of inefficient DMUs may also change from the input-oriented case.

SBM-C(V or GRS) (The slacks-based measure of efficiency under the constant (variable or general) returns-to-scale assumption.)

This code solves the SBM model. The format of the input data and the output results is the same as for the CCR case. In the GRS (general returns-to-scale) case, L (the lower bound) and U (the upper bound) of the sum of the intensity vector λ must be supplied through keyboard. The defaults are $L = 0.8$ and $U = 1.2$.

SBM-I-C(V or GRS) (The input-oriented slacks-based measure of efficiency under the constant (variable or general) returns-to-scale assumption).

This code solves the SBM model in input-orientation. Therefore, output slacks (shortfalls) are not accounted for in this efficiency measure.

SBM-O-C(V or GRS) (The output-oriented SBM model under constant (variable or general) returns-to-scale assumption).

This is the output-oriented version of the SBM model, which puts emphasis on the output shortfalls. Therefore, input slacks (excesses) are not accounted for in this efficiency measure.

WeightedSBM-(I(O))-C(V or GRS) (The weighted SBM model under constant (variable or general) returns-to-scale assumption).

This code solves the weighted SBM model introduced in Section 4.4.7. See Appendix B.5.12 for input data format.

Hybrid-C(V) (The hybrid measure of efficiency).

This model solves the Hybrid measure of efficiency under the constant (variable) returns-to-scale assumptions. See B.5.11 for the data format of this model.

FDH (The free disposal hull model). This code solves the FDH model and produces the worksheets “Score,” “Projection,” “Graph1” and “Graph2.” The data format is the same as for the CCR model.

4.12 PROBLEM SUPPLEMENT FOR CHAPTER 4

Problem 4.1

It is suggested on occasion that the weak (or Farrell) efficiency value, θ^* , be used to rank DMU performances as determined from CCR or BCC models.

Assignment : Discuss possible shortcomings in this approach.

Suggested Response : There are two main shortcomings that θ^* possesses for the proposed use. First, the value of θ^* is not a complete measure and, instead, the nonzero slacks may far outweigh the value of $(1 - \theta^*)$. Second, the θ^* values will generally be determined relative to different reference groups. Note, therefore, that a value of $\theta^* = 0.9$ for DMU A means that its purely technical efficiency is 90% of the efficiency generated by members of its (efficient) reference group. Similarly a value of $\theta^* = 0.8$ for DMU B refers to its performance relative to a different reference group. This does *not* imply that DMU B is less efficient than DMU A.

Ranking can be useful, of course, so when this is wanted recourse should be had to an explicitly stated principle of ranking. This was done by the Texas Public Utility Commission, for example, which is required by law to conduct “efficiency audits” of the 75 electric cooperatives which are subject to its jurisdiction in Texas. See Dennis Lee Thomas “Auditing the Efficiency of Regulated Companies through the Use of Data Envelopment Analysis: An Application to Electric Cooperatives,” Ph.D. Thesis, Austin, TX: University of Texas at Austin, Graduate School of Business, 1985.²³ As discussed in Thomas (subsequently Chairman of the Texas PUC), the ranking principle was based on dollarizing the inefficiencies. This included assigning dollar values to the

slack as well as to the purely technical inefficiencies. For instance, line losses in KWH were multiplied by the revenue charges at prices that could otherwise be earned from sales of delivered power. Fuel excesses, however, were costed at the purchase price per unit. Thomas then also remarks that the reference groups designated by DEA were used by PUC auditors to help in their evaluation by supplying comparisons DMUs for the cooperative being audited. See also A. Charnes, W.W. Cooper, D. Divine, T.W. Ruefli and D. Thomas "Comparisons of DEA and Existing Ratio and Regression Systems for Effecting Efficiency Evaluations of Regulated Electric Cooperatives in Texas," *Research in Government and Nonprofit Accounting* 5, 1989, pp.187-210.

Problem 4.2

Show that the CCR model is not translation invariant.

Suggested Answer : Let us examine the CCR model via a simple one input and one output case, as depicted in Figure 4.8. The efficient frontier is the line connecting the origin O and the point B . The CCR-efficiency of A is evaluated by PQ/PA . If we translate the output value by a given amount, say 2, in the manner shown in Figure 4.8, the origin will move to O' and the efficient frontier will shift to the dotted line extending beyond B from O' . Thus, the CCR-efficiency of A is evaluated, under the new coordinate system, by PR/PA , which differs from the old value PQ/PA . A similar observation of the translation of the input item shows that the CCR model is *not* invariant for translation of both *input* and *output*. From this we conclude that the condition $\sum_{j=1}^n \lambda_j = 1$ plays a crucial role which is exhibited in limited fashion in the BCC model and more generally in Additive models.

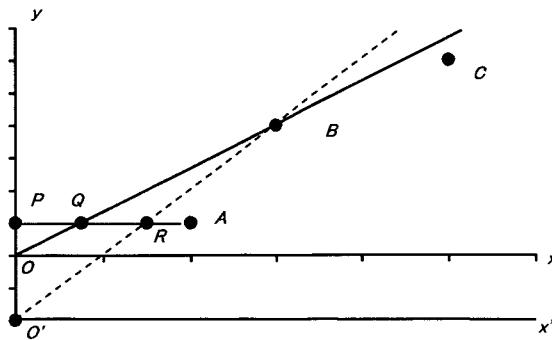


Figure 4.8. Translation in the CCR Model

Problem 4.3

As noted in the text, the critical condition for translation invariance is $e\lambda = 1$. This suggests that parts of the BCC model may be translation invariant. Therefore,

Assignment : Prove that the input-oriented (output-oriented) BCC model is translation invariant with respect to the coordinates of outputs (inputs).

Answer : The (input-oriented) BCC model is:

$$\begin{aligned} \min \quad & \theta_B \\ \text{subject to} \quad & \theta_B x_o - X\lambda - s^- = \mathbf{0} \\ & Y\lambda - s^+ = y_o \\ & e\lambda = 1 \\ & \lambda \geq \mathbf{0}, \quad s^- \geq \mathbf{0}, \quad s^+ \geq \mathbf{0}. \end{aligned} \quad (4.117)$$

If we translate the output data Y by some constants ($\beta_r : r = 1, \dots, s$) to obtain

$$y'_{rj} = y_{rj} + \beta_r, \quad (r = 1, \dots, s : j = 1, \dots, n) \quad (4.118)$$

then (4.117) becomes, in terms of the components of Y' ,

$$\sum_{j=1}^n (y'_{rj} - \beta_r) \lambda_j - s_r^+ = \sum_{j=1}^n y'_{rj} \lambda_j - \beta_r - s_r^+ = y'_{ro} - \beta_r, \quad (r = 1, \dots, s) \quad (4.119)$$

so subtracting β_r from both sides gives

$$\sum_{j=1}^n y'_{rj} \lambda_j - s_r^+ = y'_{ro}, \quad (r = 1, \dots, s) \quad (4.120)$$

which is the same as the original expression for this output constraint. Notice that the convexity condition $e\lambda = 1$ is used to derive the above relation. Notice also that this result extends to inputs *only* in the special case of $\theta_B^* = 1$.

It should now be readily apparent that if $(\theta_B^*, \lambda^*, s^{-*}, s^{+*})$ is an optimal solution of the original BCC model, then it is also optimal for the translated problem.

By the same reasoning used for the input-oriented case, we conclude that the output-oriented BCC model is translation invariant with respect to the input data X but *not* to the output data.

Comment and References : The property of translation invariance for Additive models was first demonstrated in Ali and Seiford (1990), as noted in the text. It has been carried further in J.T Pastor (1996).²⁴ See also R.M. Thrall (1996)²⁵ who shows that this property does not carry over to the dual of Additive models.

Problem 4.4

Statement : Consider the following version of an Additive model

$$\max \quad \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \quad (4.121)$$

$$\begin{aligned} \text{subject to } & \sum_{j=1}^n \widehat{x}_{ij} \lambda_j + s_i^- = \widehat{x}_{io}, \quad i = 1, \dots, m \\ & \sum_{j=1}^n \widehat{y}_{rj} \lambda_j - s_r^+ = \widehat{y}_{ro}, \quad r = 1, \dots, s \end{aligned}$$

where the variables are also constrained to be nonnegative and the symbol “ $\widehat{\cdot}$ ” means that the data are stated in natural logarithmic units. (Unlike (ADD_o) , as given in (4.34)-(4.38), we have omitted the condition $e\lambda = 1$ because this restrict results to the case of constant returns to scale.)

Assignment : Part I: Use anti-logs (base e) to derive yet another DEA model called the “multiplicative model,” and relate the resulting formulation to the Cobb-Douglas forms of production frontiers that have played prominent role in many econometric and statistical studies.

Part II: Relate the conditions for full (100%) efficiency to Definition 5.1 for the Additive model.

Suggested Response: : Part I: Let an optimal solution be represented by $\lambda_j^*, s_i^{-*}, s_r^{+*}$. Then, taking anti-logs of (4.121) we obtain

$$\begin{aligned} x_{io} &= \prod_{j=1}^n x_{ij}^{\lambda_j^*} e^{s_i^{-*}} = a_i^* \prod_{j=1}^n x_{ij}^{\lambda_j^*}, \quad i = 1, \dots, m \\ y_{ro} &= \prod_{j=1}^n y_{rj}^{\lambda_j^*} e^{-s_r^{+*}} = b_r^* \prod_{j=1}^n y_{rj}^{\lambda_j^*}, \quad r = 1, \dots, s \end{aligned} \quad (4.122)$$

where $a_i^* = e^{s_i^{-*}}$, $b_r^* = e^{-s_r^{+*}}$. This shows that each x_{io} and y_{ro} is to be considered as having been generated by Cobb-Douglas (=log-linear) processes with estimated parameters indicated by the starred values of the variables.

Part II: To relate these results to ratio forms we proceed in a manner analogous to what was done for (4.7)-(4.10) and turn to the dual (multiplier) form of (4.121),

$$\begin{aligned} \min & \quad \sum_{i=1}^m v_i \widehat{x}_{io} - \sum_{r=1}^s u_r \widehat{y}_{ro} \\ \text{subject to } & \quad \sum_{i=1}^m v_i \widehat{x}_{ij} - \sum_{r=1}^s u_r \widehat{y}_{rj} \geq 0, \quad j = 1, \dots, n \\ & \quad v_i \geq 1, \quad i = 1, \dots, m \\ & \quad u_r \geq 1, \quad r = 1, \dots, s. \end{aligned} \quad (4.123)$$

This transformation to logarithm values gives a result that is the same as in the ordinary Additive model. Hence, no new software is required to solve this multiplicative model. To put this into an easily recognized efficiency evaluation

form we apply anti-logs to (4.123) and obtain

$$\begin{aligned} \max \quad & \prod_{r=1}^s y_{ro}^{u_r} / \prod_{i=1}^m x_{io}^{v_i} \\ \text{subject to} \quad & \prod_{r=1}^s y_{rj}^{u_r} / \prod_{i=1}^m x_{ij}^{v_i} \leq 1, \quad j = 1, \dots, n \\ & v_i, u_r \geq 1, \forall i, r. \end{aligned} \quad (4.124)$$

To obtain conditions for efficiency we apply antilogs to (4.121) and use the constraints in (4.123) to obtain

$$\max \frac{\prod_{r=1}^s e^{s_r^+}}{\prod_{i=1}^m e^{-s_i^-}} = \frac{\prod_{r=1}^s \prod_{j=1}^n y_{rj}^{\lambda_j^*} / y_{ro}}{\prod_{i=1}^m \prod_{j=1}^n x_{ij}^{\lambda_j^*} / x_{io}} \geq 1. \quad (4.125)$$

The lower bound represented by the unity value on the right is obtainable if and only if all slacks are zero. Thus, the conditions for efficiency of the multiplicative model are the same as for the Additive model where the latter has been stated in logarithmic units. See Definition 5.1. We might also note that the expression on the left in (4.125) is simpler and easier to interpret and the computation via (4.121) is straightforward with the \hat{x}_{ij} and \hat{y}_{rj} stated in logarithmic units.

In conclusion we might note that (4.125) is not units invariant unless $\sum_{j=1}^n \lambda_j^* = 1$. This “constant-returns-to-scale condition” can, however, be imposed on (4.121) if desired. See also A. Charnes, W.W. Cooper, L.M. Seiford and J. Stutz “Invariant Multiplicative Efficiency and Piecewise Cobb-Douglas Envelopments” *Operations Research Letters* 2, 1983, pp.101-103.

Comments : The class of multiplicative models has not seen much used in DEA applications. It does have potential for uses in extending returns-to-scale treatments to situations that are not restricted to concave cap functions like those which are assumed to hold in the next chapter. See R.D. Banker and A. Maindiratta “Piecewise Loglinear Estimation of Efficient Production Surfaces” *Management Science* 32, 1986, pp.126-135. Its properties suggest possible uses not only in its own right but also in combination with other models and, as will be seen in the next chapter, it can be used to obtain scale elasticities that are not obtained with other DEA models.

Problem 4.5

Work out the dual version of the weighted SBM model in Section 4.4.8 and interpret the results.

Suggested Response:

$$\max \quad \mathbf{u} \mathbf{y}_o - \mathbf{v} \mathbf{x}_o \quad (4.126)$$

$$\text{subject to} \quad -\mathbf{v}X + \mathbf{u}Y \leq \mathbf{0} \quad (4.127)$$

$$v_i x_{io} \geq \frac{1}{m} w_i^- \quad (i = 1, \dots, m) \quad (4.128)$$

$$u_r y_{ro} \geq \frac{1 - \mathbf{v}x_o + \mathbf{u}y_o}{s} w_r^+ \quad (r = 1, \dots, s) \quad (4.129)$$

Thus, each $v_i x_{io}$ is bounded below by a value proportional to $1/m$ times the selected weight w_i^- for input and the value of unity increment for the y_o and decrement for the x_o combined in the numerator on s , the number of outputs in the denominator.

Problem 4.6

Prove Theorem 4.12.

Suggested Response: Let an optimal solution of the (*Hybrid*) for (\bar{x}_o, \bar{y}_o) be $(\theta^{**}, \phi^{**}, \lambda^{**}, s^{NR-**}, s^{NR+**})$. It holds that

$$\theta^{**} \bar{x}_o^R = \theta^{**} \theta^* x_o^R \geq X^R \lambda^{**} \quad (4.130)$$

$$\bar{x}_o^{NR} = X^{NR} \lambda^{**} + s^{NR-**} \quad (4.131)$$

$$\phi^{**} \bar{y}_o^R = \phi^{**} \phi^* y_o^R \geq Y^R \lambda^{**} \quad (4.132)$$

$$\bar{y}_o^{NR} = Y^{NR} \lambda^{**} - s^{NR+**} \quad (4.133)$$

$$\theta^{**} \leq 1, \quad \phi^{**} \geq 1. \quad (4.134)$$

From (4.131) and (4.133) we have:

$$x_o^{NR} = X^{NR} \lambda^{**} + (s^{NR-**} + s^{NR-*}) \quad (4.135)$$

$$y_o^{NR} = Y^{NR} \lambda^{**} - (s^{NR+**} + s^{NR+*}), \quad (4.136)$$

where * refers to an optimum of the originally stated problem and ** refers to the optimum of the problem secured by projecting these solutions to the efficient frontier. We now show that $\theta^{**} = \phi^{**} = 1$ and all slacks are zero.

From the above expressions, we can see that $(\theta^{**} \theta^*, \phi^{**} \phi^*, \lambda^{**}, s^{NR-**} + s^{NR-*}, s^{NR+**} + s^{NR+*})$ is a feasible solution of the (*Hybrid*) for (x_o, y_o) . From the optimality of $(\theta^*, \phi^*, \lambda^*, s^{NR-*}, s^{NR+*})$, we have the inequality:

$$\begin{aligned} & \frac{1 - \frac{m_1}{m}(1 - \theta^{**} \theta^*) - \frac{1}{m} \sum_{i=1}^{m_2} (s_i^{NR-**} + s_i^{NR-*}) / x_{io}^{NR}}{1 + \frac{s_1}{s} (\phi^{**} \phi^* - 1) + \frac{1}{s} \sum_{r=1}^{s_2} (s_r^{NR+**} + s_r^{NR+*}) / y_{ro}^{NR}} \\ & \geq \frac{1 - \frac{m_1}{m}(1 - \theta^*) - \frac{1}{m} \sum_{i=1}^{m_2} s_i^{NR-*} / x_{io}^{NR}}{1 + \frac{s_1}{s} (\phi^* - 1) + \frac{1}{s} \sum_{r=1}^{s_2} s_r^{NR+*} / y_{ro}^{NR}}. \end{aligned} \quad (4.137)$$

Noting that $\theta^{**} \leq 1, \phi^{**} \geq 1, s^{NR-**} \geq 0, s^{NR+**} \geq 0$, this inequality holds if and only if we have:

$$\theta^{**} = 1, \quad \phi^{**} = 1, \quad s^{NR-**} = 0, \quad s^{NR+**} = 0. \quad (4.138)$$

Thus, the projected DMU (\bar{x}_o, \bar{y}_o) is hybrid-efficient.

Problem 4.7

Develop the input (output)-oriented Hybrid model by neglecting the denominator (numerator) of the objective function (4.92) of (*Hybrid*).

Suggested Response:

$$\begin{aligned}
 (Hybrid - I) \quad \rho_I^* &= \min \quad 1 - \frac{m_1}{m}(1 - \theta) - \frac{1}{m} \sum_{i=1}^{m_2} s_i^{NR-} / x_{io}^{NR} \\
 \text{subject to} \quad &\theta x_o^R \geq X^R \lambda \\
 &x_o^{NR} = X^{NR} \lambda + s^{NR-} \\
 &y_o^R \leq Y^R \lambda \\
 &y_o^{NR} \leq Y^{NR} \lambda \\
 &\theta \leq 1, \lambda \geq 0, s^{NR-} \geq 0. \\
 (Hybrid - O) \quad \rho_O^* &= \min \frac{1}{1 + \frac{s_1}{s}(\phi - 1) + \frac{1}{s} \sum_{r=1}^{s_2} s_r^{NR+} / y_{ro}^{NR}} \\
 \text{subject to} \quad &x_o^R \geq X^R \lambda \\
 &x_o^{NR} \geq X^{NR} \lambda \\
 &\phi y_o^R \leq Y^R \lambda \\
 &y_o^{NR} = Y^{NR} \lambda - s^{NR+} \\
 &\phi \geq 1, \lambda \geq 0, s^{NR+} \geq 0.
 \end{aligned} \tag{4.139}$$

This leads to:

$$\rho_I^* \geq \rho^* \text{ and } \rho_O^* \geq \rho^*. \tag{4.140}$$

Problem 4.8

Referring to the Hybrid model, observe a single radial input (output) case and show that this reduces to a non-radial model.

Suggested Response: Let us assume that the number of the radial inputs (outputs) is one, i.e., $m_1 = 1$ ($s_1 = 1$). Then the first constraint of (*Hybrid*) becomes:

$$\theta x_o^R \geq X^R \lambda.$$

We can transform this single (not vector) inequality into:

$$x_o^R \geq X^R \lambda + (1 - \theta)x_o^R. \tag{4.141}$$

Let us define $s^{R-} = (1 - \theta)x_o^R$ (≥ 0). Then, the θ -related term in the objective function of (*Hybrid*) becomes:

$$-\frac{1}{m}s^{R-} / x_o^R.$$

Thus, the model is reduced to the non-radial case. The inequality in (4.141) is binding at the optimal solution.

Similarly, the single radial output case can be reduced to a non-radial case.

However, the single non-radial input (output) case cannot be reduced to a radial model.

Problem 4.9

Prove Lemma 4.2.

Suggested Response: In the (CCR) model, let us define

$$\pi = \theta/\phi. \quad (4.142)$$

Then (CCR) becomes:

$$\begin{aligned} (CCR) \quad & \rho_{CCR}^* = \min \pi \\ \text{subject to} \quad & \pi x_o \geq X(\lambda/\phi) \\ & y_o \leq Y(\lambda/\phi) \\ & \phi \geq 1, \lambda \geq 0. \end{aligned} \quad (4.143)$$

Essentially this is equivalent to (CCR-I). Similarly, we can demonstrate that $\rho_{CCR}^* = \rho_{CCR-O}^*$.

Problem 4.10

Prove Theorem 4.13.

Suggested Response: We will demonstrate the inequality (4.109), since others can be derived similarly. Let an optimal solution of (CCR) be $(\theta^*, \phi^*, \lambda^*)$. Then, after partition into radial and non-radial parts, we have:

$$\begin{aligned} \theta^* x_o^R &\geq X^R \lambda^* \\ \theta^* x_o^{NR} &\geq X^{NR} \lambda^* \\ \phi^* y_o^R &\leq Y^R \lambda^* \\ \phi^* y_o^{NR} &\leq Y^{NR} \lambda^*. \end{aligned}$$

Let us define

$$s^{NR-} = (1 - \theta^*)x_o^{NR} (\geq 0), \quad s^{NR+} = (\phi^* - 1)y_o^{NR} (\geq 0).$$

Then, $(\theta^*, \phi^*, \lambda^*, s^{NR-}, s^{NR+})$ is a feasible solution for (Hybrid). Thus, we have $\rho_{CCR}^* \geq \rho^*$ showing the first inequality of (4.109).

Let an optimal solution of (Hybrid) be $(\theta^*, \phi^*, \lambda^*, s^{NR-*}, s^{NR+*})$ and define s^{R-*} and s^{R+*} as:

$$s^{R-*} = (1 - \theta^*)x_o^R (\geq 0), \quad s^{R+*} = (\phi^* - 1)y_o^R (\geq 0).$$

Then we have:

$$x_o^R = X^R \lambda^* + s^{R-*}, \quad y_o^R = Y^R \lambda^* - s^{R+*}.$$

Hence, after changing the radial part into non-radial and defining $s^{-*} = (s^{R-*}, s^{NR-*})$, $s^{+*} = (s^{R+*}, s^{NR+*})$, the obtained $(\lambda^*, s^{-*}, s^{+*})$ is a feasible solution for (SBM). Thus we have $\rho^* \geq \rho_{SBM}^*$.

Problem 4.11

Assignment: Suppose the objective for ERM in (4.110) is changed from its fractional form to the following

$$\max \frac{\sum_{r=1}^s \phi_r}{s} - \frac{\sum_{i=1}^m \theta_i}{m}. \quad (4.144)$$

Using the relations in (4.112) transform this objective to the objective for the corresponding Additive model and discuss the properties of the two measures.

Suggested Response: Transforming (4.144) in the suggested manner produces the following objective for the corresponding Additive model.

$$\max z = \frac{\sum_{r=1}^s \left(1 + \frac{s_r^+}{y_{ro}}\right)}{s} - \frac{\sum_{i=1}^m \left(1 - \frac{s_i^-}{x_{io}}\right)}{m} = \sum_{r=1}^s \frac{s_r^+}{y_{ro}} + \sum_{i=1}^m \frac{s_i^-}{x_{io}}. \quad (4.145)$$

The ERM constraints are similarly transformed to the constraints of an Additive model as described in the proof of Theorem 4.15.

Discussion: The objective in (4.144) jointly maximizes the “inefficiencies” measured by the ϕ_r and minimizes the “efficiencies” measured by the θ_i . This is analogous to the joint minimization represented by the objective in (4.110) for ERM but its z^* values, while non-negative, may exceed unity. The objective stated on the right in (4.145) jointly maximizes the “inefficiencies” in both the s_r^+/y_{ro} and s_i^-/x_{io} . This is analogous to the joint maximization of both of these measures of inefficiency in SBM except that, again, its value may exceed unity.

Problem 4.12

Solve the hospital example in Table 1.5 of Chapter 1 using the SBM-C code in the DEA-Solver and decompose inefficiency into each input/output factor.

Suggested Answer: The worksheet “Decomposition” gives the results exhibited in Table 4.7. From this table, we can see that DMU C’s score $(0.829) = (1-0.048)/(1+0.149)$. This indicates the shortage in inpatient dominates others in inefficiency.

Notes

1. R. D. Banker, A. Charnes and W.W. Cooper (1984) “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis,” *Management Science* 30, pp.1078-1092
2. SBM was first proposed by K. Tone (1997), “A Slacks-based Measure of Efficiency in Data Envelopment Analysis,” Research Reports, Graduate School of Policy Science, Saitama University and subsequently published in *European Journal of Operational Research* 130 (2001), pp.498-509.

Table 4.7. Decomposition of Efficiency Score

DMU	Score	Input Excess Doctor Inefficiency	Input Excess Nurse Inefficiency	Output Shortage Outpatient Inefficiency	Output Shortage Inpatient Inefficiency
A	1	0	0	0	0
B	1	0	0	0	0
C	0.829	0.048	0	0	0.149
D	1	0	0	0	0
E	0.729	0	0.011	0.357	0
F	0.688	0.178	0	0	0.194
G	0.880	0.003	0	0	0.133
H	0.773	0.053	0	0	0.224
I	0.904	0	0.054	0	0.046
J	0.768	0.126	0	0	0.138
K	0.865	0.098	0	0.043	0
L	0.933	0	0.019	0	0.051

3. For a discussion of this constant-returns-to-scale Additive model see A.I. Ali and L.M. Seiford, "The Mathematical Programming Approach to Efficiency Measurement," in *The Measurement of Productive Efficiency: Techniques and Applications*, H. Fried, C. A. Knox Lovell, and S. Schmidt (editors), Oxford University Press, London, (1993).

4. This distance is measured in what is called an l_1 -metric for the slack vectors s^- and s^+ . Also called the "city-block" metric, the objective in (4.34) is directed to maximizing the distance as measured by the sum of the input plus output slacks.

5. T. Ahn, A. Charnes and W.W. Cooper (1988) "Efficiency Characterizations in Different DEA Models," *Socio-Economic Planning Sciences*, 22, pp.253-257.

6. P.L. Brockett, W.W. Cooper, L.L. Golden, J.J. Rousseau and Y. Wang "DEA Evaluation of Organizational Forms and Distribution System in the U.S. Property and Liability Insurance Industry," *International Journal of Systems Science*, 1998.

7. A.I. Ali and L.M. Seiford (1990) "Translation Invariance in Data Envelopment Analysis," *Operations Research Letters* 9, pp.403-405.

8. This term is taken from the literature on dimensional analysis. See pp.123-125, in R.M. Thrall "Duality, Classification and Slacks in DEA" *Annals of Operations Research* 66, 1996.

9. See R.S. Färe and C.A.K. Lovell (1978) "Measuring the Technical Efficiency of Production," *Journal of Economic Theory* 19, pp.150-162. See also R.R. Russell (1985) "Measures of Technical Efficiency," *Journal of Economic Theory* 35, pp.109-126.

10. "Hybrid" is proposed in K. Tone (2004), "A Hybrid Measure of Efficiency in DEA," GRIPS Research Report Series I-2004-0003.

11. See the Note 9 references.

12. J.T. Pastor, J.L. Ruiz and I. Sirvent (1999), "An Enhanced DEA Russell Graph Efficiency Measure," *European Journal of Operational Research* 115, pp.596-607.

13. R.S. Färe, S. Grosskopf and C.A.K Lovell (1985), *The Measurement of Efficiency of Production* Boston: Kluwer-Nijhoff.

14.I. Bardhan, W.F. Bowlin, W.W. Cooper and T. Sueyoshi (1996), "Models and Measures for Efficiency Dominance in DEA. Part II: Free Disposal Hull (FDH) and Russell Measure (RM) Approaches" *Journal of the Operations Research Society of Japan* 39, pp.333-344.

15. See the paper by T. Ahn and L.M. Seiford "Sensitivity of DEA Results to Models and Variable Sets in a Hypothesis Test Setting: The Efficiency of University Operations" in Y. Ijiri, ed., *Creative and Innovative Approaches to the Science of Management* (New York: Quorum Books, 1993). In this paper Ahn and Seiford find that U.S. public universities are more efficient than private universities when student outputs are emphasized but the reverse is true when research is emphasized. The result, moreover, is found to be invariant over all of the DEA models used in this study. For "methods" cross-checking, C.A.K. Lovell, L. Walters and L. Wood compare statistical regressions with DEA results in "Stratified Models of Education Production Using Modified DEA and Regression Analysis," in *Data Envelopment Analysis: Theory, Methodology and Applications* (Norwell, Mass.: Kluwer Academic Publishers, 1994.)

16. See Chapter 5 for detail.

17. J.T. Pastor (1996), "Translation Invariance in DEA: A Generalization," *Annals of Operations Research* 66, pp.93-102.

18. R.M. Thrall (1996) "The Lack of Invariance of Optimal Dual Solutions Under Translation," *Annals of Operations Research* 66, pp.103-108.

19. Deprins D., L. Simar and H. Tulkens (1984), "Measuring Labor Efficiency in Post Offices," in M. Marchand, P. Pestieau and H. Tulkens, eds. *The Performance of Public Enterprises: Concepts and Measurement* (Amsterdam, North Holland), pp.243-267.

20. See the review on pp. 205-210 in C.A.K. Lovell "Linear Programming Approaches to the Measurement and Analysis of Productive Efficiency," *TOP* 2, 1994, pp. 175-243.

21. Bowlin, W.F., J. Brennan, A. Charnes, W.W. Cooper and T. Sueyoshi (1984), "A Model for Measuring Amounts of Efficiency Dominance," Research Report, The University of Texas at Austin, Graduate School of Business. See also Bardhan I., W.F. Bowlin, W.W. Cooper and T. Sueyoshi "Models and Measures for Efficiency Dominance in DEA," *Journal of the Operations Research Society of Japan* 39, 1996, pp.322-332.

22. R.M. Thrall (1999), "What is the Economic Meaning of FDH?" *Journal of Productivity Analysis*, 11, pp.243-250.

23. Also available in microfilm form from University Microfilms, Inc., Ann Arbor, Michigan.

24. See the Note 18 reference.

25. See the Note 19 reference.

5 RETURNS TO SCALE

5.1 INTRODUCTION

This chapter deals with returns to scale along with modifications and extensions that are needed to use these concepts in actual applications. We start with conceptual formulations obtained from economics for which we use Figure 5.1. The function $y = f(x)$ in the upper portion of the diagram is a “production function” defined so that the value of y is maximal for every x . This definition assumes that technical efficiency is always achieved. Therefore, points like P , which lie inside the production possibility set, do not find a place in the concept that we are presently considering. Only points on the frontier defined by $y = f(x)$ are of interest.

The lower diagram portrays the behavior of average productivity ($a.p. = y/x$) and marginal productivity ($m.p. = dy/dx$) which are obtained from the upper diagram by noting that y/x corresponds to the slope of the ray from the origin to y and dy/dx is the derivative of $f(x)$ at this same point.

As can be seen, the slopes of the rays increase with x until x_0 is reached after which the slopes of the corresponding rays (and hence $a.p.$) begin to decrease. In a similarly visualized manner, the derivative $dy/dx = m.p.$ increases with x until the inflection point at $f(x_1)$ is reached after which it begins to decrease. As shown in the lower diagram, $m.p.$ lies above $a.p.$ to the left of x_0 and below

it to the right. This means that output is changing proportionally faster than input to the left of x_0 while the reverse situation occurs to the right of x_0 .

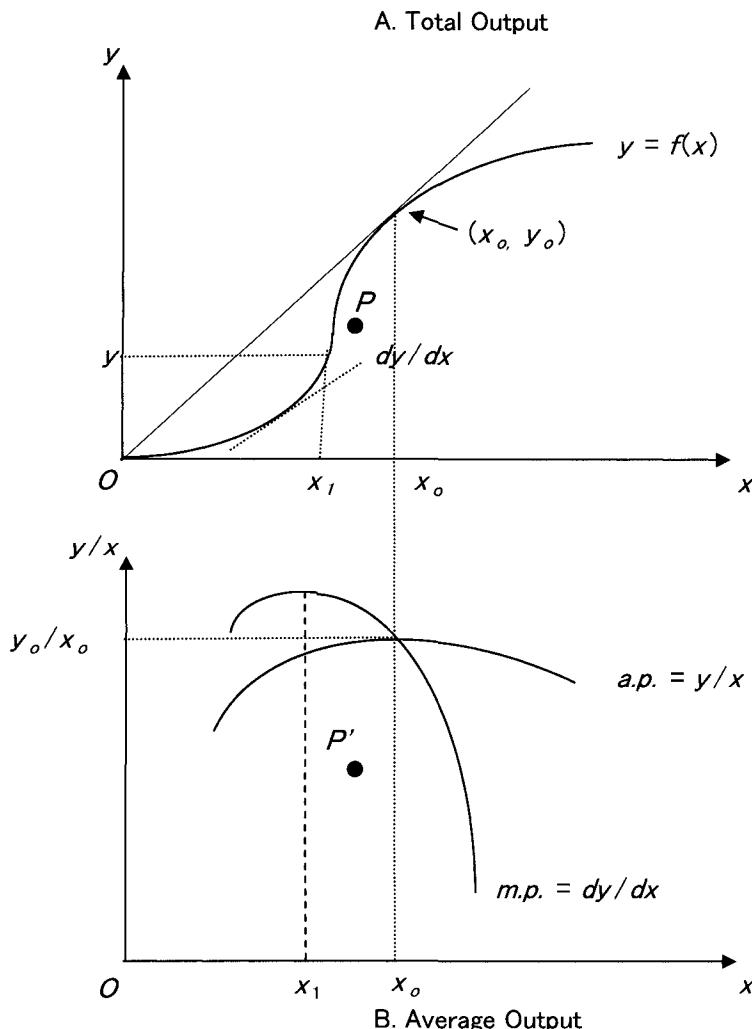


Figure 5.1. Returns to Scale

We can formalize this more precisely by determining where the maximum value of $a.p.$ occurs by differentiating $a.p.$ with respect to x via

$$\frac{d(y/x)}{dx} = \frac{x dy/dx - y}{x^2} = 0,$$

which at $x = x_0$ gives

$$e(x) = \frac{x}{y} \frac{dy}{dx} = \frac{d \ln y}{d \ln x} = 1. \quad (5.1)$$

Here $e(x)$ is defined as a logarithmic derivative to reflect the fact that proportional changes are being considered. Called an “elasticity” in economics it measures the *relative* change in output compared to the *relative* change in input. At x_0 , where *a.p.* is maximal, we have $e(x) = 1$ so that $dy/dx = y/x$ and the rate of change of output to input does not change.

Stated differently, $e(x) = 1$ represents constant returns-to-scale because the proportional increase in y given by dy/y is the same as the proportional increase in x given by dx/x . To the left of x_0 we will have $dy/dx > y/x$ so $e(x) > 1$ and returns to scale is increasing. To the right of x_0 we have $dy/dx < y/x$ so $e(x) < 1$ and returns to scale is decreasing with $dy/y < dx/x$.

We now extend the above example to the case of multiple inputs while retaining the assumption that output is a scalar represented by the variable y . To avoid confusing changes in scale with changes in mix, we write

$$y = f(\theta x_1, \theta x_2, \dots, \theta x_m)$$

and use the scalar θ to represent increase in scale when $\theta > 1$. Thus, keeping input mix constant, — we increase all input by the common scale θ — we can represent “elasticity of scale” by

$$\varepsilon(\theta) = \frac{\theta}{y} \frac{dy}{d\theta}. \quad (5.2)$$

Then, at least locally, we have increasing returns-to-scale when $\varepsilon(\theta) > 1$ and decreasing returns-to-scale when $\varepsilon(\theta) < 1$. Finally $\varepsilon(\theta) = 1$ when returns to scale is constant. All analyses are thus conducted from a point (x_1, \dots, x_m) on the efficient frontier without changing the input proportion. That is, the mix is held constant.

To treat realistic applications we will need to modify these formulations. For instance, to deal with multiple outputs and inputs simultaneously we need to replace the “maximal value of y ” which characterizes the production frontier in the single output case. Because this “maximum output” becomes ambiguous in multiple output situations we have recourse to the Pareto-Koopmans definition of efficiency given in Chapter 3. We also need to allow for occurrences of points like P and P' in Figure 5.1 which are not on the frontier. We must therefore develop methods that enable us to distinguish adjustments that eliminate such inefficiencies and adjustments associated with returns-to-scale inefficiencies. That is, we will need to be able to distinguish increasing inputs that achieve more (or less) than proportional increase in outputs at points on the efficient frontier and separate them from output increases resulting from the elimination of Pareto-Koopmans inefficiencies.

It is possible that such more-than-proportional increases in outputs are *less* valuable than the input increases required to obtain them. In economics this issue is solved by recourse to cost, profit or revenue functions — such as we treat later in this text. However, in many cases the unit price or cost data needed for such analyses are not available. A knowledge of returns to scale is worthwhile in any case and the means for attaining this knowledge by DEA occupies the rest of this chapter.

5.2 GEOMETRIC PORTRAYALS IN DEA

Figure 5.2 is a modification of Figure 5.1 to provide a simplified version of situations likely to be encountered in actual applications. To start with some of the differences, we note that A in Figure 5.2 represents the smallest of the observed input values and thereby reflects the fact that it is usually necessary to commit at least a minimum of resources — i.e., a positive threshold input value must be achieved — before any output is observed.

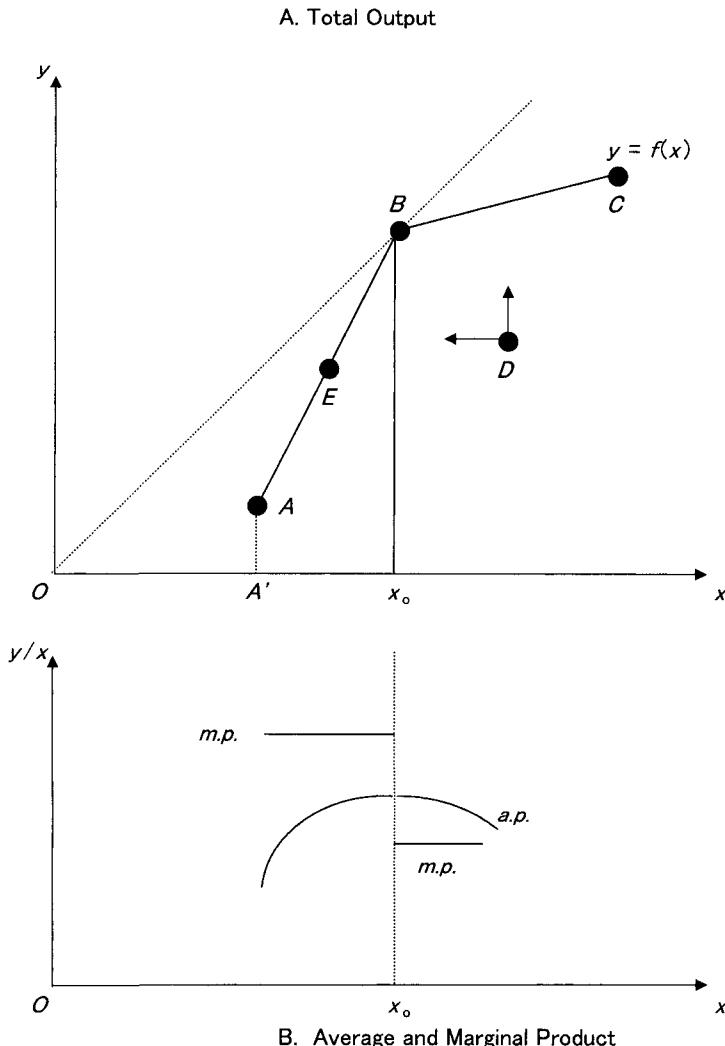


Figure 5.2. Returns to Scale: One Output-One Input

The smooth (everywhere differentiable) function used in Figure 5.1 now gives way to a piecewise linear function shown at the top of Figure 5.2, with conse-

quences like the failure of *m.p.* to intersect *a.p.* at its maximum point in the bottom figure. The failure of the derivative to exist at x_0 is there reflected in a jump associated with the “staircase-like form” that occurs at the gap exhibited for *m.p.* in Figure 5.2.

The concept of a tangent plane (or line) for dy/dx in Figure 5.1 now needs to be extended to the more general concept of a “supporting hyperplane” which we develop as follows. A hyperplane H_o in an $(m+s)$ dimensional input-output space passing through the point represented by the vectors $(\mathbf{x}_o, \mathbf{y}_o)$ can be expressed by the equation:

$$H_o : \quad \mathbf{u}(\mathbf{y} - \mathbf{y}_o) - \mathbf{v}(\mathbf{x} - \mathbf{x}_o) = 0, \quad (5.3)$$

where $\mathbf{u} \in R^s$ and $\mathbf{v} \in R^m$ are coefficient vectors. Now we define u_0 , a scalar, by

$$u_0 = \mathbf{u}\mathbf{y}_o - \mathbf{v}\mathbf{x}_o. \quad (5.4)$$

Thus, the hyperplane in (5.3) can be expressed as:

$$\mathbf{u}\mathbf{y} - \mathbf{v}\mathbf{x} - u_0 = 0. \quad (5.5)$$

Generally, a hyperplane divides the space into two halfspaces. If the hyperplane H_o contains the production possibility set P in only one of the halfspaces, then it is said to be a *supporting hyperplane* of P at the point $(\mathbf{x}_o, \mathbf{y}_o)$. That is, a supporting hyperplane touches the production possibility set at this point.

More concretely, for every $(\mathbf{x}, \mathbf{y}) \in P$, associated with any DMU,

$$\mathbf{u}\mathbf{y} - \mathbf{v}\mathbf{x} - u_0 \leq 0 \quad (5.6)$$

holds. From the definition of P for the BCC model and the above property of a supporting hyperplane, it can be demonstrated that

$$\mathbf{u} \geq \mathbf{0}, \quad \mathbf{v} \geq \mathbf{0}, \quad (5.7)$$

because the production possibility set is on only one side of the hyperplane.

In addition, a linear equation is invariant under multiplication by a nonzero number, so, to eliminate this source of indeterminacy, we add the constraint

$$\mathbf{v}\mathbf{x}_o = 1, \quad (5.8)$$

which we refer to as a *normalization* condition. P is generated by the data set represented in the matrices (X, Y) and hence (5.6) can be replaced by

$$\mathbf{u}Y - \mathbf{v}X - u_0 \mathbf{e} \leq \mathbf{0}, \quad (5.9)$$

which represents this relation for all DMUs. Also, from (5.4) and (5.8),

$$\mathbf{u}\mathbf{y}_o - u_0 = 1. \quad (5.10)$$

Thus, it is concluded that the vector of coefficients (v, u, u_0) of the supporting hyperplane H_o is a (dual) optimal solution of the BCC model. Hence we have,

Theorem 5.1 *If a DMU (x_o, y_o) is BCC-efficient, the normalized coefficient (v, u, u_0) of the supporting hyperplane to P at (x_o, y_o) gives an optimal solution of the BCC model and vice versa.*

We will make the sense (and use) of this theorem more concrete in the next section. Here we simply pause to note that the supporting hyperplane at a point like B in Figure 5.2 is not unique. The support property applies when the hyperplane (here a line) coincides with the line from A to B . It continues to apply, however, as the line is rotated around B until coincidence with the line from B to C is achieved. Thus the uniqueness associated with the assumption that $f(x)$ is analytic, as in Figure 5.1, is lost. This lack of uniqueness needs to be reflected in the returns-to-scale characterizations that we now supply.

5.3 BCC RETURNS TO SCALE

In (4.7)-(4.10) of the preceding chapter, the dual (multiplier) problem of the BCC model was expressed as:¹

$$(DBCC_o) \quad \max \quad z = \mathbf{u} \mathbf{y}_o - u_0 \quad (5.11)$$

$$\text{subject to} \quad \mathbf{v} \mathbf{x}_o = 1 \quad (5.12)$$

$$-\mathbf{v} X + \mathbf{u} Y - u_0 \mathbf{e} \leq \mathbf{0} \quad (5.13)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, u_0 \text{ free in sign}, \quad (5.14)$$

where the free-in-sign variable, u_0 , is associated with the condition $e\lambda = 1$ in the corresponding envelopment model given as (BCC_o) in (4.2)-(4.6). See also Table 4.1.

When a DMU is BCC-efficient we can invoke Theorem 5.1 and employ the sign of u_0^* to portray the situation for returns to scale. This can be done in a simple manner by returning to A in the picture portrayed in the top diagram of Figure 5.2. It is apparent that returns to scale is increasing at A because the slopes of the rays going from the origin to $f(x)$ are increasing at this point. The supports for all rotations at A starting with coincidence at $A'A$ and continuing toward the line AB will all have negative intercepts which we can associate with $u_0^* < 0$. At E the support which has this same negative intercept property is unique. Thus, all of the supports have this same negative intercept property — which, we might note, is also shared by the tangent line at (x_0, y_0) in Figure 5.1. Jumping over to C in Figure 5.2 we note that the supports at this point will all have positive intercepts which are associated with $u_0^* > 0$. All such supports can thus be associated with decreasing returns to scale at C . Finally, at B we note that the rotations include the support represented by the broken line through the origin which we can associate with $u_0^* = 0$ as the intercept where returns to scale is constant — as was the case in Figure 5.1 where the tangent line also goes through the origin.

Using $(\mathbf{x}_o, \mathbf{y}_o)$ to represent a point with coordinate values corresponding to the multiple inputs and outputs recorded for DMU_o, we formalize the above in the following theorem for use with the Model (BCC_o) — from Banker and Thrall (1992)² —

Theorem 5.2 *Assuming that $(\mathbf{x}_o, \mathbf{y}_o)$ is on the efficient frontier the following conditions identify the situation for returns to scale at this point,*

- (i) *Increasing returns-to-scale prevails at $(\mathbf{x}_o, \mathbf{y}_o)$ if and only if $u_0^* < 0$ for all optimal solutions.*
- (ii) *Decreasing returns-to-scale prevails at $(\mathbf{x}_o, \mathbf{y}_o)$ if and only if $u_0^* > 0$ for all optimal solutions.*
- (iii) *Constant returns-to-scale prevails at $(\mathbf{x}_o, \mathbf{y}_o)$ if and only if $u_0^* = 0$ in any optimal solution.*

Chasing down *all* optimal solutions can be onerous. We can avoid this task, however, and also eliminate the efficiency assumption that is embedded in this theorem (so that we can treat points like D in Figure 5.2). This can be accomplished as follows. Assume that a solution is available from the BCC model represented in (4.2)-(4.6). This gives the information needed to project $(\mathbf{x}_o, \mathbf{y}_o)$ into $(\widehat{\mathbf{x}}_o, \widehat{\mathbf{y}}_o)$, a point on the efficient frontier, via

$$\begin{aligned}\widehat{x}_{io} &= \theta_B^* x_{io} - s_i^{-*}, \quad i = 1, \dots, m \\ \widehat{y}_{ro} &= y_{ro} + s_r^{+*}, \quad r = 1, \dots, s\end{aligned}\tag{5.15}$$

where $\theta_B^*, s_i^{-*}, s_r^{+*}$ are part of a solution to (4.2)-(4.6) — which identifies this as a BCC model.

From this solution to the BCC envelopment model, most computer codes report the value of u_0^* as part of the optimal solution for the dual (multiplier) problem. Suppose therefore that our solution yields $u_0^* < 0$. To take advantage of this information in deciding whether (i) or (iii) applies from the above theorem, we utilize a path opened by Banker, Bardhan and Cooper (1996)³ and replace the customary dual (multiplier) problem as given in (5.11)-(5.14) with the following,

$$\begin{array}{ll}\max & u_0 \\ \text{subject to} & -\sum_{i=1}^m v_i x_{ij} + \sum_{r=1}^s u_r y_{rj} - u_0 \leq 0, \quad j = 1, \dots, n; j \neq o \\ & -\sum_{i=1}^m v_i \widehat{x}_{io} + \sum_{r=1}^s u_r \widehat{y}_{ro} - u_0 \leq 0, \quad j = o \\ & \sum_{i=1}^m v_i \widehat{x}_{io} = 1, \quad \sum_{r=1}^s u_r \widehat{y}_{ro} - u_0 = 1 \\ & u_0 \leq 0 \\ & v_i \geq 0, \quad u_r \geq 0\end{array}\tag{5.16}$$

where the \hat{x}_{io} and \hat{y}_{ro} are obtained from (5.15), and the added constraint $\sum_{r=1}^s u_r \hat{y}_{ro} - u_0 = 1$ is necessary for the solution to lie on the efficient frontier. Complemented by Definition 2.1, the condition becomes sufficient. Because (\hat{x}_o, \hat{y}_o) is on the efficiency frontier for the BCC model with (5.15) there is no need to introduce the non-Archimedean element $\varepsilon > 0$.

The constraint $u_0 \leq 0$ means that the maximum of (5.16) cannot exceed zero. If the latter value is achieved the returns to scale is constant by (iii) of Theorem 5.2. If it is not achieved – i.e., the maximum is achieved with $u_0^* < 0$ – then condition (i) is applicable and returns to scale is decreasing.

The formulation (5.16) therefore forms part of a two-stage solution procedure in which one first solves (5.11)-(5.14) to obtain a solution with $u_0^* < 0$. This is then followed by the use of (5.16) to see whether an alternative optimum is present which can achieve $u_0^* = 0$. This corresponds to the treatment we accorded to D in Figure 5.2 where (a) one first eliminated the technical and mix inefficiencies and then (b) determined whether output increases (or decreases) could be obtained which are more, less, or exactly proportional to input variations that might then be undertaken along the efficient frontier.

We have formulated (5.16) for use when $u_0^* < 0$ occurred in the initial stage. If $u_0^* > 0$ had occurred then the inequality $u_0 \leq 0$ would be replaced by $u_0 \geq 0$ and the objective in (5.16) would be reoriented to “ $\min u_0$.” (This will be illustrated by a numerical example in the next section.)

5.4 CCR RETURNS TO SCALE

In keeping with the DEA literature, we have used “constant returns-to-scale” to characterize the CCR model. This is technically correct but somewhat misleading because this model can also be used to determine whether returns to scale is increasing or decreasing. This is accomplished by slightly sharpening the conditions discussed in Section 5.3 to obtain the disjunction in the following theorem by Banker and Thrall (1992).

Theorem 5.3 *Let (x_o, y_o) be a point on the efficient frontier. Employing a CCR model in envelopment form to obtain an optimal solution $(\lambda_1^*, \dots, \lambda_n^*)$, returns to scale at this point can be determined from the following conditions,*

- (i) *If $\sum_{j=1}^n \lambda_j^* = 1$ in any alternate optimum then constant returns-to-scale prevails.*
- (ii) *If $\sum_{j=1}^n \lambda_j^* > 1$ for all alternate optima then decreasing returns-to-scale prevails.*
- (iii) *If $\sum_{j=1}^n \lambda_j^* < 1$ for all alternate optima then increasing returns-to-scale prevails.*

Banker and Thrall (1992) proved this theorem on the assumption that (x_o, y_o) is on the efficient frontier. However, we now utilize the route provided by Banker, Chang and Cooper (1996)⁴ to eliminate the need for this assumption. This is accomplished by proceeding in a two-stage manner analogous to the

preceding section. That is, we utilize the information available from the first stage solution to formulate the following problem,

$$\begin{aligned} \min \quad & \sum_{j=1}^n \hat{\lambda}_j - \varepsilon \left(\sum_{i=1}^m \hat{s}_i^- + \sum_{r=1}^s \hat{s}_r^+ \right) \\ \text{subject to} \quad & \theta_o^* \mathbf{x}_o = \sum_{j=1}^n \mathbf{x}_j \hat{\lambda}_j + \hat{\mathbf{s}}^- \\ & \mathbf{y}_o = \sum_{j=1}^n \mathbf{y}_j \hat{\lambda}_j - \hat{\mathbf{s}}^+ \\ & 1 \leq \sum_{j=1}^n \hat{\lambda}_j \\ & 0 \leq \hat{\lambda}_j, \quad \forall j, \end{aligned} \quad (5.17)$$

where $\mathbf{x}_j, \mathbf{y}_j$ are vectors of input and output values with components corresponding to $\mathbf{x}_o, \mathbf{y}_o$ representing the observation for the DMU_o to be evaluated. The slack vectors $\mathbf{s}^-, \mathbf{s}^+$ as well as the vector $\hat{\lambda}$ are constrained to be nonnegative.

Here θ_o^* is obtained from the first stage solution, as in (5.15). However, we do not utilize the slack values, as is done in (5.15), because these are automatically provided by the way the objective is formulated in (5.17). Thus,

$$\begin{aligned} \theta_o^* \mathbf{x}_o - \hat{\mathbf{s}}^- * &= \sum_{j=1}^n \mathbf{x}_j \hat{\lambda}_j^* \\ \mathbf{y}_o + \hat{\mathbf{s}}^+ * &= \sum_{j=1}^n \mathbf{y}_j \hat{\lambda}_j^* \end{aligned} \quad (5.18)$$

exhibits the coordinates of an efficient point obtained from the expressions on the left in terms of a nonnegative combination of data from the other DMUs on the right. As noted in Definition 3.4, the DMUs on the right for which $\hat{\lambda}_j^* > 0$ constitute efficient members of the reference set used to evaluate DMU_o. The efficient point represented on the left of (5.18) is therefore expressed in terms of the DMUs that define this portion (a facet) of the efficient frontier.

We have formulated (5.17) for use after a first stage solution with $\sum_{j=1}^n \lambda_j^* > 1$ has been achieved. The case where $\sum_{j=1}^n \lambda_j^* = 1$ needs no further treatment since by (i) of Theorem 5.3 constant returns to scale will then prevail at the efficient point represented in (5.18). The remaining case — viz., $\sum_{j=1}^n \lambda_j^* <$

$1 -$ is addressed by writing $\sum_{j=1}^n \hat{\lambda}_j \leq 1$ and replacing the objective with
 $\max \quad \sum_{j=1}^n \hat{\lambda}_j + \varepsilon(\sum_{i=1}^m \hat{s}_i^- + \sum_{r=1}^s \hat{s}_r^+).$

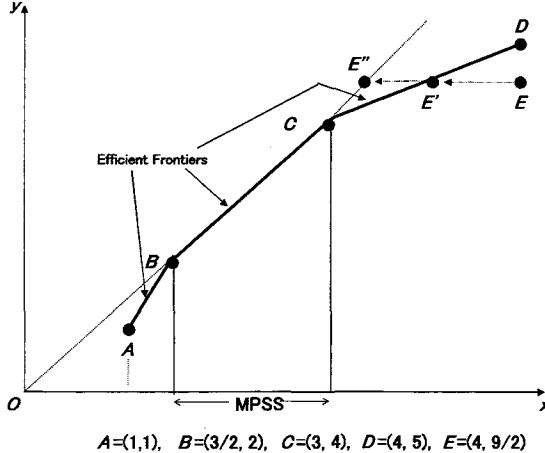


Figure 5.3. Most Productive Scale Size

For a numerical illustration using (5.17) we start with E in Figure 5.3. With coordinates $(x, y) = (4, 9/2)$ it is the only point that is not on the efficient frontier for either the BCC model (represented by the solid lines connecting A, B, C and D) or the efficient frontier of the CCR model (represented by the dotted line that passes through points B and C). In this case the first stage solution yields $\theta_o^* = 27/32$ with an alternate optimum for either $\lambda_B^* = 9/4$ or $\lambda_C^* = 9/8$ and all other variables = 0. Hence in both optima we have $\sum_{j=1}^n \lambda_j^* > 1$. To check whether condition (i) or (ii) in Theorem 5.3 is applicable we apply (5.17) to the (x, y) values located at the bottom of Figure 5.3, to obtain

$$\begin{aligned} \min \quad & (\hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E) - \varepsilon(\hat{s}^- + \hat{s}^+) \\ \text{subject to} \quad & \theta_o^* x_o = 27/8 = 1\hat{\lambda}_A + 3/2\hat{\lambda}_B + 3\hat{\lambda}_C + 4\hat{\lambda}_D + 4\hat{\lambda}_E + \hat{s}^- \\ & y_o = 9/2 = 1\hat{\lambda}_A + 2\hat{\lambda}_B + 4\hat{\lambda}_C + 5\hat{\lambda}_D + 9/2\hat{\lambda}_E - \hat{s}^+ \\ & 1 \leq \hat{\lambda}_A + \hat{\lambda}_B + \hat{\lambda}_C + \hat{\lambda}_D + \hat{\lambda}_E \\ & 0 \leq \hat{\lambda}_A, \hat{\lambda}_B, \hat{\lambda}_C, \hat{\lambda}_D, \hat{\lambda}_E, \hat{s}^-, \hat{s}^+. \end{aligned}$$

This has an optimum at $\hat{\lambda}_B^* = 9/4$ with all other variables equal to zero. Because this is the minimizing solution, it follows that $\sum_{j=1}^n \hat{\lambda}_j^* > 1$ in all alternate optima so (ii) – as the applicable condition in Theorem 5.3 – means that decreasing returns to scale prevails.

To illustrate the opposite case we turn to A , a point on the BCC-efficient frontier in Figure 5.3 where returns to scale is increasing. For this case we report a first stage solution with $\min \theta_o = \theta_o^* = 3/4$ which is associated with

alternate optima consisting of $\lambda_B^* = 1/2$ or $\lambda_C^* = 1/4$, and all other variables zero. To deal with this case, for which $\sum_{j=1}^n \lambda_j^* < 1$, we modify (5.17) in the manner indicated by the discussion following (5.18) to obtain

$$\begin{aligned} \max \quad & (\widehat{\lambda}_A + \widehat{\lambda}_B + \widehat{\lambda}_C + \widehat{\lambda}_D + \widehat{\lambda}_E) + \varepsilon(\widehat{s}^- + \widehat{s}^+) \\ \text{subject to} \quad & 3/4 = 1\widehat{\lambda}_A + 3/2\widehat{\lambda}_B + 3\widehat{\lambda}_C + 4\widehat{\lambda}_D + 4\widehat{\lambda}_E + \widehat{s}^- \\ & 1 = 1\widehat{\lambda}_A + 2\widehat{\lambda}_B + 4\widehat{\lambda}_C + 5\widehat{\lambda}_D + 9/2\widehat{\lambda}_E - \widehat{s}^+ \\ & 1 \geq \widehat{\lambda}_A + \widehat{\lambda}_B + \widehat{\lambda}_C + \widehat{\lambda}_D + \widehat{\lambda}_E \\ & 0 \leq \widehat{\lambda}_A, \widehat{\lambda}_B, \widehat{\lambda}_C, \widehat{\lambda}_D, \widehat{\lambda}_E, \widehat{s}^-, \widehat{s}^+. \end{aligned}$$

This has an optimum solution with $\widehat{\lambda}_B^* = 1/2$ and all other values zero so that condition (iii) of Theorem 5.3 is applicable and returns to scale is found to be increasing at A .

We next turn to E and now evaluate it by the BCC model first with (5.11)-(5.14) and then with (5.16). This will allow us to show how to modify (5.16) to treat the case $u_0^* > 0$ because the first stage solution gives $\theta_B^* = 7/8$ with $\lambda_C^* = \lambda_D^* = 1/2$ and the corresponding multiplier values are $u_0^* = u^* = v^* = 1/4$.

Reorienting (5.16) for use with this first stage $u_0^* > 0$ replaces $\max u_0$ with $\min u_0$ and $u_0 \leq 0$ with $u_0 \geq 0$ as in the following example.

$$\begin{aligned} \min \quad & u_0 \\ \text{subject to} \quad & -v + u - u_0 \leq 0 \\ & -3/2v + 2u - u_0 \leq 0 \\ & -3v + 4u - u_0 \leq 0 \\ & -4v + 5u - u_0 \leq 0 \\ & -7/2v + 9/2u - u_0 \leq 0 \\ & 7/2v = 1 \\ & 9/2u - u_0 = 1 \\ & v, u \geq 0, u_0 \geq 0. \end{aligned}$$

A minimizing solution to this problem is $v^* = u^* = u_0^* = 2/7$. This confirms the results we earlier found with the CCR model — viz., returns to scale is decreasing at E' , the point with coordinate $(\widehat{x}_o, \widehat{y}_o) = (27/8, 9/2)$ in Figure 5.3. We now note that this coordinate is on the efficient frontier of the BCC model, but not for the CCR model. This is confirmed by noting that $\theta_B^* = 7/8 = 28/32$ for the BCC model applied to E whereas for the CCR model we get $\theta_o^* = 27/32$. Hence, a still further input contraction is needed to attain the CCR-efficiency available at E'' , even after E' is attained in Figure 5.3.

We are now in a position to interpret the two models in the following manner. The BCC model separates the analysis in a twofold manner by (i) evaluating technical (and mix) efficiency in its envelopment model and (ii) evaluating returns-to-scale efficiency in its dual (multiplier) model. The CCR model, on

the other hand, simultaneously evaluates both types of efficiency using only the envelopment form in the manner set forth in our treatment of Theorem 5.3.

The term “scale efficiency” in DEA involves recourse to constant returns-to-scale in a manner that generalizes the situation in Figure 5.1 where *a.p.* (average product) reached a maximum at (x_o, y_o) . In addition to treating multiple outputs and inputs simultaneously it is necessary in DEA to allow for cases in which constant returns-to-scale prevails over an entire region such as the very simple example of the segment stretching from *B* to *C* in Figure 5.3.

In the next section we refer to regions like this as MPSS (Most Productive Scale Size) and use this as a gauge of scale efficiency. Here we note that we can access this region by invoking the following projection formulas from Banker and Morey (1986),⁵

$$\begin{aligned}\hat{x}_{io}^* &\Leftarrow \frac{\theta_o^* x_{io} - \hat{s}_i^{-*}}{\sum_{j=1}^n \hat{\lambda}_j^*}, \quad i = 1, \dots, m \\ \hat{y}_{ro}^* &\Leftarrow \frac{y_{ro} + \hat{s}_r^{+*}}{\sum_{j=1}^n \hat{\lambda}_j^*}, \quad r = 1, \dots, s\end{aligned}\quad (5.19)$$

where the $\hat{\lambda}_j^*$, \hat{s}_i^{-*} , \hat{s}_r^{+*} are obtained from an optimal solution for (5.17). In fact the numerators of (5.19) are obtained from (5.15), and division by the denominators thus yield the expressions

$$\begin{aligned}\hat{x}_o^* &= \sum_{j=1}^n x_j \tilde{\lambda}_j^* \\ \hat{y}_o^* &= \sum_{j=1}^n y_j \tilde{\lambda}_j^* \\ \text{with } 1 &= \sum_{j=1}^n \tilde{\lambda}_j^*. \quad (5.20)\end{aligned}$$

To illustrate we return to the initial example in which *E* was evaluated with $\sum_{j=1}^n \hat{\lambda}_j^* = 9/4$. Applying (5.19) to $(x_o, y_o) = (27/8, 9/2)$ therefore gives

$$\begin{aligned}\hat{x}^* &= \frac{27}{8} \frac{4}{9} = \frac{3}{2} \\ \hat{y}^* &= \frac{9}{2} \frac{4}{9} = 2\end{aligned}$$

which are the coordinates of *B*, a point of constant returns-to-scale which is in the MPSS region of Figure 5.3.

We carry the discussion of MPSS further in the next section. Here we conclude by returning to the remark with which we began this chapter — viz., characterizing CCR as a constant returns-to-scale model is somewhat misleading. In fact use of the CCR model might be preferred to use of the BCC model for returns-to-scale analyses for reasons like the following:

First the CCR model generally involves fewer constraints and hence is likely to be slightly more efficient computationally. Second the solutions from this model directly provide more of the pertinent information in forms for use in projective formulas like (5.15) and (5.17) - to which we now add (5.18).

Relations between the CCR and BCC models are studied further in the next section. Here we conclude with the following theorem — due to Banker and Thrall (1992) — which relates the returns-to-scale characterizations from these two models as follows,

Theorem 5.4

- (i) $u_o^* > 0$ for all optimal solutions to a BCC model if and only if
 $\sum_{j=1}^n \hat{\lambda}_j^* > 1$ for all optimal solutions to the corresponding CCR model.
- (ii) $u_o^* < 0$ for all optimal solutions to a BCC model if and only if
 $\sum_{j=1}^n \hat{\lambda}_j^* < 1$ for all optimal solutions to the corresponding CCR model.
- (iii) $u_o^* = 0$ for some optimal solution to a BCC model if and only if
 $\sum_{j=1}^n \hat{\lambda}_j^* = 1$ for some optimal solution to the corresponding CCR model.

This is conveniently summarized in the following,

Corollary 5.1 (Corollary to Theorem 5.4)

- (i) $u_o^* > 0$ for all optimal solutions to a BCC model if and only if
 $\sum_{j=1}^n \lambda_j^* - 1 > 0$ in all optimal solutions to the corresponding CCR model.
- (ii) $u_o^* < 0$ for all optimal solutions to a BCC model if and only if
 $\sum_{j=1}^n \lambda_j^* - 1 < 0$ in all optimal solutions to the corresponding CCR model.
- (iii) $u_o^* = 0$ for some optimal solution to a BCC model if and only if
 $\sum_{j=1}^n \lambda_j^* - 1 = 0$ in some optimal solution to the corresponding CCR model.

5.5 MOST PRODUCTIVE SCALE SIZE

To further examine relations between BCC and CCR models we start with the following theorem due to Ahn, Charnes and Cooper (1989),⁶

Theorem 5.5 A DMU_o found to be efficient with a CCR model will also be found to be efficient with the corresponding BCC model and constant returns-to-scale prevail at DMU_o .

This theorem, which is simple to prove, is critically important for the topics covered in this section. It is also important to note that the converse is not necessarily true. As exemplified by E' in Figure 5.3, a DMU can be simultaneously characterized as efficient by a BCC model and inefficient by a CCR model with $\theta_o^* < \theta_B^*$. However, if $\theta_o^* = \theta_B^*$ then there will be at least one alternate optimum for this θ_o^* for which $\sum_{j=1}^n \lambda_j^* = 1$.

This theorem is directed to mathematical properties of the two types of models. To obtain interpretations that we can relate to our earlier discussions of returns to scale and to open possibilities for managerial uses we utilize

R.D. Banker's concept of Most Productive Scale Size (MPSS),⁷ which we initiate with the following expression

$$(\mathbf{x}_o\alpha, \mathbf{y}_o\beta). \quad (5.21)$$

Here $(\mathbf{x}_o, \mathbf{y}_o)$ are vectors with components corresponding to coordinates of the point being evaluated and (α, β) are scalars representing expansion or contraction factors according to whether $\alpha, \beta > 1$ or $\alpha, \beta < 1$ are applied to the inputs and outputs represented by these coordinates.

Our discussion in Section 1 of this chapter was confined to a single output so a scale factor, applied to all inputs, provided everything needed to study scale variations in a manner that would avoid confusing variations in mix with economies or diseconomies of scale. Here α plays the role previously assigned to θ for inputs and the scalar β plays a similar role for outputs.

We also need to allow for inequality relations in place of the equalities used in (5.1) and (5.2). We then want a measure of scale which is "dimensionless" (i.e., it does not depend on the units of measure used) so we utilize a ratio of these two scalars, each of which is dimensionless.

To accomplish these (and other) purposes we use the following model

$$\begin{aligned} \max \quad & \beta/\alpha \\ \text{subject to} \quad & \beta\mathbf{y}_o \leq \sum_{j=1}^n y_j \lambda_j \\ & \alpha\mathbf{x}_o \geq \sum_{j=1}^n x_j \lambda_j \\ & 1 = \sum_{j=1}^n \lambda_j \\ & 0 \leq \beta, \alpha. \end{aligned} \quad (5.22)$$

This is a fractional programming problem. By the procedures described in Problem 3.1 (in Chapter 3)⁸ we can transform this to an ordinary linear programming problem to obtain solutions $\hat{\mathbf{x}}_o = \alpha\mathbf{x}_o, \hat{\mathbf{y}}_o = \beta\mathbf{y}_o$ which we can associate with points which are MPSS by means of the following theorem,

Theorem 5.6 *A necessary condition for a DMU_o with output and input vectors \mathbf{y}_o and \mathbf{x}_o to be at MPSS is $\beta^*/\alpha^* = \max \beta/\alpha = 1$ in (5.22), in which case $\beta^* = \alpha^*$ and returns to scale will be constant.*

We illustrate for D the point with coordinates (4, 5) in Figure 5.3. Thus, applying (5.22) we obtain the following problem,

$$\begin{aligned} \max \quad & \beta/\alpha \\ \text{subject to} \quad & 5\beta \leq 1\lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + 9/2\lambda_E \\ & 4\alpha \geq 1\lambda_A + 3/2\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E \\ & 1 = \lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E \\ & 0 \leq \lambda_A, \dots, \lambda_E. \end{aligned}$$

This has an optimum with $\lambda_B^* = 1$ and $\beta^* = 2/5, \alpha^* = 3/8$. It also has an alternate optimum with $\lambda_C^* = 1$ and $\beta^* = 4/5, \alpha^* = 3/4$. In either case we have $\beta^*/\alpha^* = 16/15$ so, via Theorem 5.6, D is not at MPSS.

The information needed to bring D into MPSS is at hand in both cases. Thus, utilizing the first solution we have $\alpha^*x_o = 3/8 \times 4 = 3/2$ and $\beta^*y_o = 2/5 \times 5 = 2$ which yields new coordinates $(\hat{x}_o, \hat{y}_o) = (3/2, 2)$ where $(\hat{x}_o, \hat{y}_o) = (\alpha^*x_o, \beta^*y_o)$ as in (5.21). Reference to Figure 5.3 shows these values to be the coordinates of B , a point which is MPSS in the constant-returns-to-scale region of Figure 5.3. Turning next to the alternate optimum we have $\beta^*y_o = 4/5 \times 5 = 4$ and $\alpha^*x_o = 3/4 \times 4 = 3$, which are the coordinates of C , a point which is also MPSS in Figure 5.3. Indeed, every convex combination of these two solutions yields $\beta^*/\alpha^* = 16/15$ with components $(\hat{x}_o, \hat{y}_o) = (\alpha^*x_o, \beta^*y_o)$ corresponding to the coordinates of points in the region of constant-returns-to-scale in Figure 5.3 and, hence, each such point is MPSS.

To extend the relation $\alpha^*x_o = \hat{x}_o$ and $\beta^*y_o = \hat{y}_o$ to the case of multiple inputs and outputs we write these vector relations in the following form,

$$\begin{aligned}\mathbf{e}\beta^* &= \left(\frac{\hat{y}_{1o}}{y_{1o}}, \dots, \frac{\hat{y}_{ro}}{y_{ro}}, \dots, \frac{\hat{y}_{so}}{y_{so}} \right) \\ \mathbf{e}\alpha^* &= \left(\frac{\hat{x}_{1o}}{x_{1o}}, \dots, \frac{\hat{x}_{io}}{x_{io}}, \dots, \frac{\hat{x}_{mo}}{x_{mo}} \right),\end{aligned}\quad (5.23)$$

where \mathbf{e} is the vector with unity for all of its elements. To show that the proportions in all outputs and inputs are maintained at the values β^* and α^* , respectively, we simply note that $\beta^* = \hat{y}_{ro}/y_{ro} = \hat{y}_{ko}/y_{ko}$ implies $y_{ko}/y_{ro} = \hat{y}_{ko}/\hat{y}_{ro}$ and $\alpha^* = \hat{x}_{io}/x_{io} = \hat{x}_{lo}/x_{lo}$ implies $x_{io}/x_{lo} = \hat{x}_{io}/\hat{x}_{lo}$ so output and input mixes are both preserved.

We can also use these formulations to generalize the elasticity in Section 1 to the case of multiple-inputs and multiple-outputs by noting that

$$\frac{1 - \beta^*}{1 - \alpha^*} = \frac{y_{ro} - \hat{y}_{ro}}{y_{ro}} \quad / \quad \frac{x_{io} - \hat{x}_{io}}{x_{io}} = \frac{x_{io}}{y_{ro}} \frac{\Delta y_{ro}}{\Delta x_{io}}. \quad (5.24)$$

For each of these mix pairs we then have a measure of scale change associated with movement to the region of MPSS.

To see what these results mean we return to D in Figure 5.3 to note that the change from D to B can be associated with output-to-input ratios which have values $5/4 = 15/12$ at D and $2/(3/2) = 16/12$ at B . We can then use (5.21) to note that $\beta^*/\alpha^* = 16/15$ gives the change of output in ratio form when the inputs are set at $x = 12$ in both cases.

Returning to (5.24) we recall that an elasticity measures the proportional change in output relative to the proportional change in input. Hence when input decreases from $x = 4$ to $x = 3/2$ the proportional decrease is $(4 - 3/2)/4 = 5/8$ and $(1 - \beta^*)/(1 - \alpha^*) = 24/25$ which means that the proportional decrease in output will be $24/23 \times 5/8 = 3/5$ — which is the proportional output change in moving from D to B . As is to be expected in the case of decreasing returns-to-scale, the proportional decrease in output is less than the proportional decrease

in input. This is confirmed by the movement from D to C which gives

$$\frac{x_o}{y_o} \frac{\Delta y_o}{\Delta x_o} = \frac{4}{5} \cdot \frac{5-4}{4-3} = \frac{4}{5} \quad \text{or} \quad \frac{1}{5} = \frac{5-4}{5} = \frac{\Delta y_o}{y_o} < \frac{\Delta x_o}{x_o} = \frac{4-3}{4} = \frac{1}{4},$$

and the relative change in output is smaller than the relative change in input in this case as well.

The immediately preceding developments assume that the observations for the DMUs being evaluated are on the BCC-efficient frontiers. This is to be expected because, as previously emphasized, elasticity and returns to scale are both frontier concepts. However, this is not a limiting assumption because formula (5.15) may be invoked when these kinds of concepts (and their related measures) are to be utilized.

When only access to MPSS is of interest we can develop an approach from the following

Definition 5.1 (MPSS) *For DMU_o to be MPSS both of the following conditions need be satisfied: (i) $\beta^*/\alpha^* = 1$ and (ii) all slacks are zero.*

This is intended to mean that the slacks in *all* alternative optima are zero. This contingency is covered in a manner analogous to the two-stage procedure outlined in Chapter 3. That is, we employ (5.22) — or its associated linear programming problem — to obtain a solution β^*/α^* . If this has any nonzero slacks it is not necessary to proceed further. However, if all slacks are zero we check to see that no alternate optimum has nonzero slacks by moving to a second stage in which $\beta = \beta^*$ and $\alpha = \alpha^*$ are fixed and the slacks are maximized. The following formulas may then be used to obtain a point with coordinates y_{ro}^*, x_{io}^* which are MPSS,

$$\begin{aligned} \beta^* y_{ro} + s_r^{+*} &= y_{ro}^*, \quad r = 1, \dots, s \\ \alpha^* x_{io} - s_i^{-*} &= x_{io}^*, \quad i = 1, \dots, m \end{aligned} \tag{5.25}$$

This allows us to conclude this discussion on a note which is concerned with practical implementation of the concepts we have been considering. In a conference on the economics of the iron and steel industry one of the economic consultants, Harold H. Wein, complained that the returns-to-scale formulations in economics were of little use. He had never participated in a case where changes in scale were separated from changes in mix, so formulas like (5.2) were of no use either in (a) arriving at such decisions or (b) analyzing their consequences.

Dr. Wein might also have complained about the restriction to a single output and the assumed prevalence of technical efficiency. Here, however, we have addressed all of these concerns. Thus we have moved to the multiple output case and removed the assumption that technical efficiency has been achieved. We also have provided (5.25) to disentangle mix inefficiencies for both inputs and outputs (as represented in the nonzero slacks) and separated them from the scale inefficiencies represented by values of $\beta^*/\alpha^* \neq 1$. Hence we have

allowed for inequalities in inputs and in outputs and we have treated both multiple inputs and multiple outputs in a way that distinguishes scale and mix inefficiencies.

5.6 FURTHER CONSIDERATIONS

Still another approach to returns to scale analysis may be used as given in Tone (1996),⁹ which we develop as follows:

Theorem 5.7 (Tone(1996)) *In the BCC model a reference set to any BCC inefficient DMU does not include both increasing and decreasing returns-to-scale DMUs at the same time.*

Proof : Let the BCC inefficient DMU be $(\mathbf{x}_o, \mathbf{y}_o)$ and its projection onto the efficient frontier by (5.15) be $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$. Since $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ is BCC-efficient, there exists an optimal weight $(\mathbf{v}_e, \mathbf{u}_e, u_{0e})$ that satisfies

$$\mathbf{u}_e \hat{\mathbf{y}}_o - u_{0e} = 1 \quad (5.26)$$

$$\mathbf{v}_e \hat{\mathbf{x}}_o = 1 \quad (5.27)$$

$$-\mathbf{v}_e X + \mathbf{u}_e Y - u_{0e} \mathbf{e} \leq \mathbf{0} \quad (5.28)$$

$$\mathbf{v}_e > \mathbf{0}, \quad \mathbf{u}_e > \mathbf{0}. \quad (5.29)$$

Hence, for $j \in E_o$ (a reference set to $(\mathbf{x}_o, \mathbf{y}_o)$), we have:

$$-\mathbf{v}_e \mathbf{x}_j + \mathbf{u}_e \mathbf{y}_j - u_{0e} \leq 0. \quad (5.30)$$

From (5.26) and (5.27),

$$-\mathbf{v}_e \hat{\mathbf{x}}_o + \mathbf{u}_e \hat{\mathbf{y}}_o - u_{0e} = 0. \quad (5.31)$$

Since

$$\hat{\mathbf{x}}_o = \sum_{j \in E_o} \lambda_j^* \mathbf{x}_j \quad \text{and} \quad \hat{\mathbf{y}}_o = \sum_{j \in E_o} \lambda_j^* \mathbf{y}_j,$$

we obtain:

$$-\mathbf{v}_e \sum_{j \in E_o} \lambda_j^* \mathbf{x}_j + \mathbf{u}_e \sum_{j \in E_o} \lambda_j^* \mathbf{y}_j - u_{0e} = 0. \quad (5.32)$$

This equation can be transformed, using $\sum_{j \in E_o} \lambda_j^* = 1$, into:

$$\sum_{j \in E_o} \lambda_j^* (-\mathbf{v}_e \mathbf{x}_j + \mathbf{u}_e \mathbf{y}_j - u_{0e}) = 0. \quad (5.33)$$

However, from (5.30), (5.33) we have some $\lambda_j^* > 0$ ($j \in E_o$) to satisfy $\sum_{j \in E_o} \lambda_j^* = 1$. Hence, for each j we have the equation:

$$-\mathbf{v}_e \mathbf{x}_j + \mathbf{u}_e \mathbf{y}_j - u_{0e} = 0. \quad (\forall j \in E_o) \quad (5.34)$$

Thus, $(\mathbf{v}_e, \mathbf{u}_e, u_{0e})$ are coefficients of a supporting hyperplane at each $(\mathbf{x}_j, \mathbf{y}_j)$ for every $j \in E_o$. Let $t_j = 1/\mathbf{v}_e \mathbf{x}_j (> 0)$, then $t_j \mathbf{v}_e = \mathbf{v}'_e, t_j \mathbf{u}_e = \mathbf{u}'_e, t_j u_{0e} = u'_{0e}$

gives $u'_e y_j - u'_{0e} = 1$ which is an optimal solution for the multiplier form associated with $j \in E_o$.

Suppose that E_o contains an *increasing* returns-to-scale DMU a and a *decreasing* returns-to-scale DMU b . Then, u_{0e} must be negative, since returns to scale is increasing at DMU a . At the same time, u_{0e} must be positive, since DMU b has decreasing returns-to-scale characteristics. This leads to a contradiction. \square

We also have a corollary from the above,

Corollary 5.2 *Let a reference set to a BCC-inefficient DMU (x_o, y_o) be E_o . Then, E_o consists of one of the following combinations of BCC-efficient DMUs where IRS, CRS and DRS stand for increasing, constant and decreasing returns-to-scale, respectively.*

- (i) All DMUs have IRS.
- (ii) Mixture of DMUs with IRS and CRS.
- (iii) All DMUs have CRS.
- (iv) Mixture of DMUs with CRS and DRS.
- (v) All DMUs show DRS.

Based on the above observations, we have

Theorem 5.8 (Characterization of Return to Scale)

Let the BCC-projected activity of a BCC-inefficient DMU (x_o, y_o) be (\hat{x}_o, \hat{y}_o) and a reference set to (x_o, y_o) be E_o . Then, (\hat{x}_o, \hat{y}_o) belongs to

1. IRS, if E_o consists of DMUs in categories (i) or (ii) of Corollary 5.2, and
2. DRS, if E_o consists of DMUs in categories (iv) or (v).

Proof: In the case of (i) or (ii), E_o contains at least one DMU with IRS and any supporting hyperplane at (\hat{x}_o, \hat{y}_o) is also a supporting hyperplane at the IRS DMU, as shown in the proof of Theorem 5.7. Thus, the upper bound of u_{0e} must be negative. By the same reasoning, in the case of (iv) or (v), the projected activity has DRS. \square

However, in the case of (iii) we cannot identify the returns to scale of (\hat{x}_o, \hat{y}_o) directly from its reference set. In this case, (\hat{x}_o, \hat{y}_o) may display IRS, CRS or DRS. We need further LP calculations for identifying its returns-to-scale characteristics as described below.

Using the preceding results a procedure for identifying the returns-to-scale characteristics of DMU_o represented by vectors (x_o, y_o) can be developed as follows.

First we solve the envelopment form of the (input-oriented) BCC model for (x_o, y_o) via (4.2)-(4.6).

(i) If DMU_o is found to be BCC-efficient then we can, at the same time, find an optimal u_0^* of the multiplier form (5.11)-(5.14) as a value of an optimal dual variable associated with the envelopment model constraint $e\lambda = 1$.

(ia) If $u_0^* = 0$ then we can identify that *constant* returns-to-scale prevails at (x_o, y_o) by Theorem 5.2.

(ib) If $u_0^* < (>)0$ then we solve the following LP using its dual (envelopment form), since the envelopment form is computationally much easier than the multiplier form.

$$\begin{aligned} [\text{Multiplier form}] \quad \max (\min) \quad & u_0 \\ \text{subject to} \quad & -vX + uY - u_0e \leq 0 \\ & vx_o = 1, \quad uy_o - u_0 = 1 \\ & v, \quad u \geq 0, \quad u_0 \text{ free in sign.} \end{aligned} \quad (5.35)$$

$$\begin{aligned} [\text{Envelopment form}] \quad \min (\max) \quad & z = \theta - \lambda_0 \quad (\lambda_0 - \theta) \\ \text{subject to} \quad & \theta x_o \geq X\lambda \\ & \lambda_0 y_o \leq Y\lambda \\ & -e\lambda + \lambda_0 = 1 \quad (e\lambda - \lambda_0 = 1) \\ & \lambda \geq 0, \quad \theta, \quad \lambda_0 \text{ free in sign.} \end{aligned} \quad (5.36)$$

Both LPs above have the same optimal value if [Envelopment form] has a finite minimum (maximum). If [Multiplier form] has an unbounded optimal value (∞ in the max case or $-\infty$ in the min cas) the [Envelopment form] has no feasible solution. Let the optimal objective value of [Multiplier form] be \bar{u}_0^* (\underline{u}_0^*), including ∞ ($-\infty$). Then by Theorem 5.2 we have,

- (a) If $\bar{u}_0^* \geq 0$ ($\underline{u}_0^* \leq 0$) we can conclude that *constant* returns-to-scale prevails at (x_o, y_o) .
- (b) If $\bar{u}_0^* < 0$ ($\underline{u}_0^* > 0$) we can conclude that *increasing* (*decreasing*) returns-to-scale prevails at (x_o, y_o) .

(ii) If DMU_o is found to be BCC inefficient then we can relate the returns-to-scale characteristics of its projected DMU (\hat{x}_o, \hat{y}_o) to those of its reference set as represented in (5.15) in the cases (iia) and (iib) in the following manner:

- (iia) The projected DMU (\hat{x}_o, \hat{y}_o) displays *increasing* returns-to-scale if the reference set of (x_o, y_o) consists of either *increasing* returns-to-scale DMUs or a mixture of *increasing* and *constant* returns-to-scale DMUs.
- (iib) The projected DMU (\hat{x}_o, \hat{y}_o) displays *decreasing* returns-to-scale if the reference set of (x_o, y_o) consists of either *decreasing* returns-to-scale DMUs or a mixture of *decreasing* and *constant* returns-to-scale DMUs.
- (iic) If the reference set of (\hat{x}_o, \hat{y}_o) consists of DMUs all belonging to *constant* returns to scale, then we apply the above procedure (i) to (\hat{x}_o, \hat{y}_o) and identify its returns to scale characteristics.

A merit of the above procedure exists in the fact that no extra computations are needed for identifying returns-to-scale characteristics of BCC inefficient DMUs in the cases (iia) and (iib). This will reduce fairly amount of computations. Numerical examples are presented in Problem 5.4.

5.7 RELAXATION OF THE CONVEXITY CONDITION

We can extend the BCC envelopment model by relaxing the convexity condition $e\lambda = 1$ to

$$L \leq e\lambda \leq U, \quad (5.37)$$

where L ($0 \leq L \leq 1$) and U ($1 \leq U$) are upper and lower bounds for the sum of the λ_j . Notice that $L = 0$ and $U = \infty$ correspond to the CCR model and $L = U = 1$ corresponds to the BCC model. Some typical extensions are described below.

1. The Increasing Returns-to-Scale (IRS) Model

The case $L = 1$, $U = \infty$ is called the *Increasing Returns-to-Scale* (IRS) or *Non-Decreasing Returns-to-Scale* (NDRS) model. The constraint on λ is :

$$e\lambda \geq 1. \quad (5.38)$$

The condition $L = 1$ assumes that we cannot reduce the scale of DMU but it is possible to expand the scale to infinity. Figure 5.4 shows the production possibility set of the IRS model for the single input and single output case.

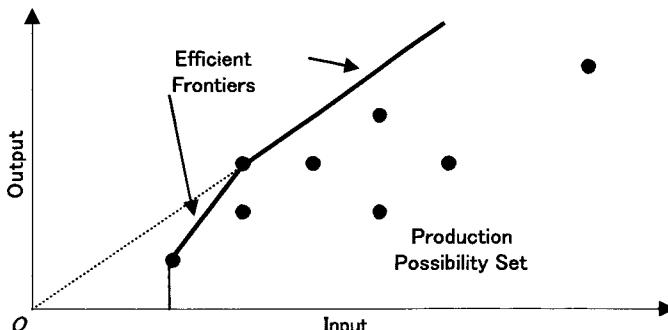


Figure 5.4. The IRS Model

The output/input ratio for any point on the efficient frontier is not decreasing with respect to input and the term NDRS is derived from this fact. That is, a proportional increase in output is always at least as great as the related proportional increase in input. In mathematical terms, we always have

$$\frac{\Delta y/y}{\Delta x/x} = \frac{x}{y} \frac{\Delta y}{\Delta x} \geq 1$$

so that

$$\frac{\Delta y}{y} \geq \frac{\Delta x}{x}$$

where Δy , Δx are the increases to be made from a frontier point with coordinate (x, y) . This model focuses on the scale efficiencies of relatively small DMUs.

2. The Decreasing Returns-to-Scale (DRS) Model

The case $L = 0$, $U = 1$ is called the *Decreasing Returns-to-Scale* (DRS) or *Non-Increasing Returns-to-Scale* (NIRS) model. The constraints on λ are:

$$0 \leq e\lambda \leq 1. \quad (5.39)$$

By the condition $U = 1$, scaling up of DMUs is interdicted, while scaling down is permitted. Figure 5.5 depicts the production possibility set. The output/input ratio of efficient frontier points is decreasing with respect to the input scale. That is, $\Delta y/y \leq \Delta x/x$ with equality (=constant returns-to-scale) for the first segment on the frontier and strict inequality holding thereafter, so this model puts the emphasis on larger DMUs where returns to scale is decreasing.

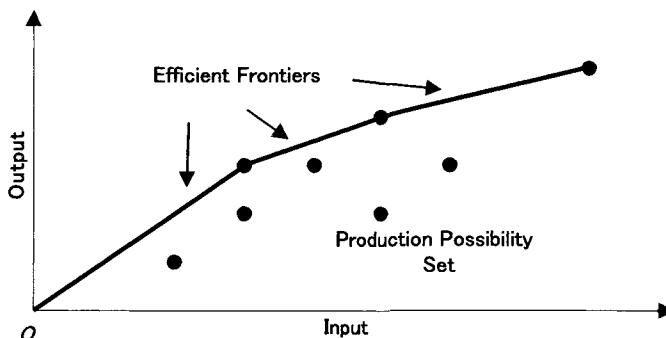


Figure 5.5. The DRS Model

3. The Generalized Returns-to-Scale (GRS) Model

The case $0 \leq L \leq 1$, $U \geq 1$ is called the *Generalized Returns-to-Scale* (GRS) model because it is possible to use these values to control the admissible range allowed for returns to scale. For instance $L = 0.8$ and $U = 1.2$ means that returns to scale can be decreasing at most in the proportion $L = 0.8$ and they can be increasing at most in the proportion $U = 1.2$. Figure 5.6 shows the production possibility set of the GRS model for the single-input and single-

output case, which sharpens the BCC model. Efficiency measured by this model is not better than that measured by BCC.

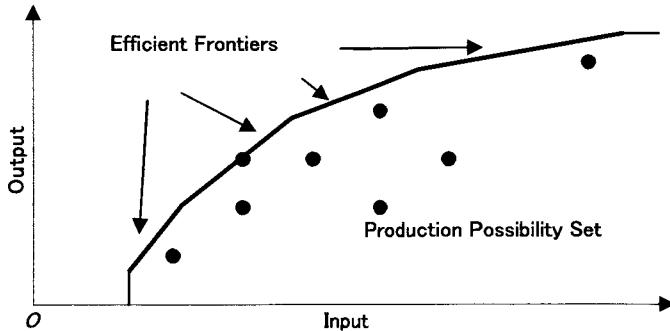


Figure 5.6. The GRS Model

It is logically true that for every DMU we have the relations

$$\theta_{CCR}^* \leq \theta_{IRS}^*, \theta_{DRS}^*, \theta_{GRS}^* \leq \theta_{BCC}^*.$$

5.8 DECOMPOSITION OF TECHNICAL EFFICIENCY

It is interesting to investigate the sources of inefficiency that a DMU might have. Are they caused by the inefficient operation of the DMU itself or by the disadvantageous conditions under which the DMU is operating?

For this purpose, comparisons of the (input-oriented) CCR and BCC scores deserve consideration. The CCR model assumes the constant returns-to-scale production possibility set, i.e., it is postulated that the radial expansion and reduction of all observed DMUs and their nonnegative combinations are possible and hence the CCR score is called *global technical efficiency*. On the other hand, the BCC model assumes that convex combinations of the observed DMUs form the production possibility set and the BCC score is called *local pure technical efficiency*. If a DMU is fully efficient (100%) in both the CCR and BCC scores, it is operating in the *most productive scale size* as pointed out in Section 5.5. If a DMU has full BCC efficiency but a low CCR score, then it is operating locally efficiently but not globally efficiently due to the scale size of the DMU. Thus, it is reasonable to characterize the *scale efficiency* of a DMU by the ratio of the two scores.

5.8.1 Scale Efficiency

Here we extend this approach to show how to decompose these inefficiencies into their component parts. Based on the CCR and BCC scores, we define *scale efficiency* as follows:

Definition 5.2 (Scale Efficiency) Let the CCR and BCC scores of a DMU be θ_{CCR}^* and θ_{BCC}^* , respectively. The scale efficiency is defined by

$$SE = \frac{\theta_{CCR}^*}{\theta_{BCC}^*}. \quad (5.40)$$

SE is not greater than one. For a BCC-efficient DMU with CRS characteristics, i.e., in the most productive scale size, its scale efficiency is one. The CCR score is called the (global) *technical efficiency* (TE), since it takes no account of scale effect as distinguished from PTE. On the other hand, BCC expresses the (local) *pure technical efficiency* (PTE) under variable returns-to-scale circumstances. Using these concepts, relationship (5.40) demonstrates a decomposition of efficiency as $\theta_{CCR}^* = \theta_{BCC}^* \times SE$ or,

$$[\text{Technical Eff. (TE)}] = [\text{Pure Technical Eff. (PTE)}] \times [\text{Scale Eff. (SE)}]. \quad (5.41)$$

This decomposition, which is unique, depicts the sources of inefficiency, i.e., whether it is caused by inefficient operation (PTE) or by disadvantageous conditions displayed by the scale efficiency (SE) or by both.

In the single input and single output case, the scale efficiency can be illustrated by Figure 5.7. For the BCC-efficient A with IRS, its scale efficiency is given by

$$SE(A) = \theta_{CCR}^*(A) = \frac{LM}{LA} < 1,$$

which denotes that A is operating locally efficient ($PTE=1$) and its overall inefficiency (TE) is caused by its failure to achieve scale efficiency (SE) as represented by LM/LA . For DMUs B and C , their scale efficiency is one, i.e.,

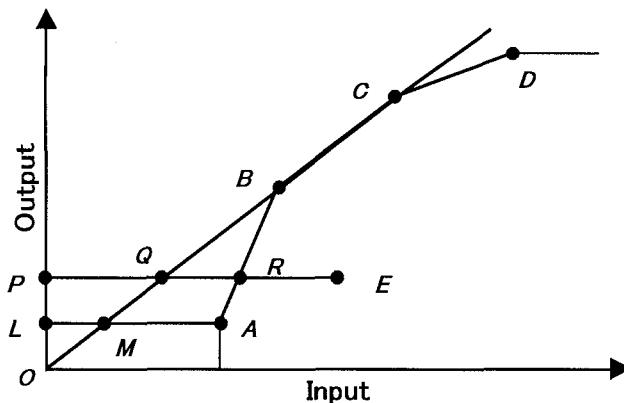


Figure 5.7. Scale Efficiency

they are operating at the most productive scale size. Their technical efficiency

is also one so they are both scale and technically efficient for both the CCR and BCC models. For the BCC-inefficient DMU E , we have

$$\text{SE}(E) = \frac{PQ}{PE} \frac{PE}{PR} = \frac{PQ}{PR},$$

which is equal to the scale efficiency of the input-oriented BCC projection R . The decomposition of E is

$$\begin{aligned}\text{TE}(E) &= \text{PTE}(E) \times \text{SE}(E) \quad \text{or} \\ \frac{PQ}{PE} &= \frac{PR}{PE} \frac{PQ}{PR}.\end{aligned}$$

Thus, E 's overall inefficiency is caused by the technically inefficient operation of E and at the same time by the disadvantageous scale condition of E measured by PQ/PR .

Although the above scale efficiency is input-oriented, we can define the *output-oriented scale efficiency* using the output-oriented scores, as well.

5.8.2 Mix Efficiency

There are two types of efficiency measures, i.e., radial and non-radial. In Section 4.4 of Chapter 4, we introduced a non-radial measure of efficiency called SBM (slacks-based measure of efficiency). In order to turn SBM to an input-orientation, we modify the objective function (4.51) to deal only with the input slacks s^- . Thus we have the input-oriented SBM model formulated as

$$\begin{aligned}(SBM_{in}) \quad \min \quad \rho_{in} &= 1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{io} \\ \text{subject to} \quad \mathbf{x}_o &= X\boldsymbol{\lambda} + \mathbf{s}^- \\ \mathbf{y}_o &= Y\boldsymbol{\lambda} - \mathbf{s}^+ \\ \boldsymbol{\lambda} &\geq \mathbf{0}, \quad \mathbf{s}^- \geq \mathbf{0}, \quad \mathbf{s}^+ \geq \mathbf{0}.\end{aligned}\tag{5.42}$$

Let an optimal solution of (SBM_{in}) be $(\rho_{in}^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$. Then we have the relationship $\rho_{in}^* \leq \theta_{CCR}^*$ (see also Problem 4.5). Furthermore, we have the following theorem.

Theorem 5.9 *The equality $\rho_{in}^* = \theta_{CCR}^*$ holds if and only if the input-oriented CCR model has zero input-slacks for every optimal solution.*

Thus, the strict inequality $\rho_{in} < \theta_{CCR}^*$ holds if and only if the CCR solution reveals an input mix inefficiency. These observations lead us to the following definition of “mix efficiency.”

Definition 5.3 (Mix Efficiency) *Let the input-oriented CCR and SBM scores of DMU_o be θ_{CCR}^* and ρ_{in}^* , respectively. The mix efficiency is defined by*

$$\text{MIX} = \frac{\rho_{in}^*}{\theta_{CCR}^*}.\tag{5.43}$$

MIX is not greater than one and we have a decomposition of the non-radial efficiency into radial and mix efficiencies as

$$[\text{Input-oriented SBM}] = [\text{Radial Eff. (TE)}] \times [\text{Mix Eff. (MIX)}]. \quad (5.44)$$

Using (5.41) (the decomposition of TE), we have the decomposition of the non-radial technical efficiency ρ_{in}^* into MIX, pure technical efficiency (PTE) and scale efficiency (SE) as

$$\rho_{in}^* = [\text{MIX}] \times [\text{PTE}] \times [\text{SE}]. \quad (5.45)$$

The above decomposition is unique and will help to interpret sources of inefficiencies for each non-radial inefficient DMU.

5.8.3 An Example of Decomposition of Technical Efficiency

We applied the above decomposition to the hospital example in Table 1.5 in Chapter 1. The results are exhibited in Table 5.1. It is observed that F 's

Table 5.1. Decomposition of Technical Efficiency

DMU	SBM ρ_{in}^*	CCR TE	BCC PTE	Mix Eff MIX	Scale Eff SE
<i>A</i>	1	1	1	1	1
<i>B</i>	1	1	1	1	1
<i>C</i>	0.852	0.883	0.896	0.965	0.985
<i>D</i>	1	1	1	1	1
<i>E</i>	0.756	0.763	0.882	0.99	0.866
<i>F</i>	0.704	0.835	0.939	0.843	0.889
<i>G</i>	0.895	0.902	1	0.992	0.902
<i>H</i>	0.774	0.796	0.799	0.972	0.997
<i>I</i>	0.905	0.96	0.989	0.942	0.971
<i>J</i>	0.781	0.871	1	0.896	0.871
<i>K</i>	0.866	0.955	1	0.907	0.955
<i>L</i>	0.936	0.958	1	0.977	0.958
Average	0.872	0.910	0.959	0.957	0.949

low SBM(0.704) is caused by MIX(0.843) and SE(0.889), and *E*'s SBM(0.756) is mainly attributed to PTE(0.882) and MIX(0.866). Although *H* is roughly efficient with respect to MIX(0.972) and SE(0.997), its low PTE(0.799) forces it to be inefficient overall with SBM=0.774. These are main causes of their inefficiency, which is measured by the non-radial slacks-based model.

5.9 AN EXAMPLE OF RETURNS TO SCALE USING A BANK MERGER SIMULATION

Since the collapse of the bubble economy in 1990, Japan has been suffering from worsening recessions. Financial institutions may be one of the most affected of all industries. Although once regarded as the most stable and profitable sector, financial institutions in Japan have been greatly weakened. During the 1980s, banks and insurance companies loaned considerable amounts of money to builders and construction companies. However, the prices of land and buildings dropped by 70-90% from their highest levels in the 1980s. In order to protect against the ruin of the financial system, the Japanese Government began to assist these institutions by using public money to prevent them from becoming insolvent. At the same time, it was recommended that these institutions should restructure their organizations, with mergers being one method of restructuring. This section is concerned with this subject, i.e., how to check whether the merger policy is effective in the case of local banks.

5.9.1 *Background*

Ordinary Japanese banks can be classified roughly into three categories: “city,” “regional” and “trust.” City banks, of which there were nine in 1997, serve almost all the 47 prefectures in Japan as well as overseas. Regional banks, numbering more than one hundred, offer limited services to customers, mainly in one prefecture, with their branches concentrated more densely within the prefecture than the city banks.

Although each regional bank is small in scale compared with an average city bank, their total shares in deposits and loans are almost at the same levels as those of city banks. Hence, their robustness is one of the main concerns of bureaucrats, politicians and financial authorities in the Government. In this section we compare a set of regional banks with nine city banks with respect to efficiencies and returns to scale. Then we simulate the effects of merging in the case of two regional banks that are hypothetically merged into one under restructuring by reducing the number of branches and employees.

5.9.2 *Efficiencies and Returns to Scale*

Table 5.2 exhibits data for nineteen banks in terms of the number of branches and employees, assets and operating net profits computed as an average over three years; 1995, 1996 and 1997. In this table the first eleven refer to regional banks and the last nine to city banks as noted by the symbols R and C, respectively. From the 104 regional banks, we chose the eleven in the northern part of Japan as typical rural area banks.

We employed 3 inputs (the number of branches and employees and assets) and one output (operating net profits) for evaluating the relative efficiency of the banks listed in the table. The CCR, BCC and Scale efficiencies and

Table 5.2. Data of 11 Regional and 9 City Banks*

No.	Bank	Branch	Employee	Assets**	Profits**
R1	Aomori	112	1,894	2,024,205	13,553
R2	Michinoku	111	1,727	1,702,017	16,540
R3	Iwate	123	1,843	2,048,014	14,760
R4	Tohoku	56	792	600,796	3,782
R5	Kita-Nippon	86	1,422	1,173,778	8,700
R6	Akita	111	1,841	2,290,274	18,590
R7	Hokuto	105	1,633	1,361,498	9,699
R8	Shonai	64	889	668,236	5,605
R9	Yamagata	90	1,611	1,563,930	9,993
R10	Yamagatashiawase	69	1,036	654,338	3,525
R11	Shokusan	73	1,011	656,794	3,177
C1	Dai-ichi Kangyo	407	17,837	53,438,938	411,149
C2	Sakura	529	18,805	52,535,262	302,901
C3	Fuji	363	15,188	51,368,976	371,364
C4	Tokyo Mitsubishi	419	20,235	78,170,694	504,422
C5	Asahi	436	13,149	29,531,193	170,835
C6	Sanwa	395	13,998	52,999,340	399,398
C7	Sumitomo	397	15,710	56,468,571	353,531
C8	Daiwa	218	8,671	16,665,573	112,121
C9	Tokai	305	11,546	31,376,180	186,640

* Source: *Analysis of Financial Statements of All Banks*,

Federation of Bankers Associations of Japan. 1995, 1996, 1997.

** Unit = One million yen \approx U.S.\$ 8300.

returns-to-scale characteristics of each bank are listed in Table 5.3. We used the input-oriented models in measuring efficiency.

The CCR results, listed in Column 2 of Table 5.3, show that the regional banks performed worse than the city banks when evaluated on the constant returns-to-scale assumption associated with this model, as evidenced by the fact that 8 out of 11 regional banks were below the average compared with 3 out of 9 city banks. However, regional bank R2 is one of 4 best performers, and furthermore it is the bank most frequently referenced for evaluating inefficient banks. It is used as a reference for all inefficient regional banks and serves as the most influential referent, i.e., with the largest λ value. It performs similarly for inefficient city banks. R2 is also the most efficient one in CCR measure. For confirmation, we might note that this bank is famous for its unique managerial strategies under the strong leadership of its owner.

The BCC scores provide efficiency evaluations using a local measure of scale, i.e. under variable returns-to-scale. In this model, two regional banks, R4 and R8, are accorded efficient status in addition to the four CCR efficient banks —

Table 5.3. Efficiencies and Returns to Scale

Bank No.	Score	CCR		BCC		Scale Score
		Reference	Score	RTS		
R1	0.714	R2 C1	0.764	Incr	0.935	
R2	1		1	Const	1	
R3	0.768	R2 C6	0.8	Incr	0.96	
R4	0.648	R2	1	Incr	0.648	
R5	0.763	R2	0.854	Incr	0.893	
R6	0.89	R2 C1 C6	0.913	Incr	0.977	
R7	0.733	R2	0.777	Incr	0.943	
R8	0.863	R2	1	Incr	0.863	
R9	0.676	R2 C1	0.777	Incr	0.87	
R10	0.554	R2	0.918	Incr	0.603	
R11	0.500	R2	0.915	Incr	0.544	
C1	1		1	Const	1	
C2	0.742	R2 C1 C6	0.747	Const	0.993	
C3	0.981	C1 C4 C6	0.99	Incr	0.991	
C4	1		1	Const	1	
C5	0.728	R2 C1	0.743	Const	0.98	
C6	1		1	Const	1	
C7	0.854	C1 C4 C6	0.866	Incr	0.986	
C8	0.853	R2 C1	0.854	Const	0.999	
C9	0.766	R2 C1	0.766	Incr	1	
Average	0.802		0.884		0.909	

which retain their previous efficient status. R4's full efficiency with the BCC model is caused by its use of the smallest amount of inputs (refer to Theorem 4.3) even though it is the lowest in the CCR score. The BCC scores do not discriminate well between regional and city banks. The regional banks exhibit scores in which 5 out of 11 are below average, while the city banks have 5 out of 9. This means that both categories are positioned, on average, at about the same relative distance from the efficient BCC frontiers, although there are differences in efficiency within each category. Both groups are performing in a similar way with regard to efficiency under the managerial situations to which they belong. The scale efficiency as defined by the ratio, CCR/BCC (see Definition 5.2 and discussions following it), exhibits large differences between the two groups. All city banks are above average whereas 6 regional banks are below it. This may mean that banks in the city group are in an advantageous condition compared with those in the regional group, and their global inefficiencies (CCR) are mainly attributed to their inefficient operations or management. The cases of the regional banks are mixed. For example, R1 has a low BCC

score and a relatively high scale efficiency among the group, meaning that the overall inefficiency (0.714) in the CCR column of R1 is caused by inefficient operations (0.764) rather than scale inefficiency (0.935). R4 has a fully efficient BCC score and a low scale efficiency (0.648). This can be interpreted to mean that the global inefficiency of this bank under CCR score is mainly attributed to disadvantageous regional conditions.

Now we turn to the returns to scale, as identified by the input-oriented BCC model and displayed in Table 5.3 under the heading "RTS." As was claimed in Theorem 5.5, banks with full efficiency in the CCR score are also efficient in the BCC model, and are in the MPSS, the region where constant returns-to-scale prevails. Banks R2, C1, C4, C6 and C9 have this status, while C2, C3 and C8 are close. Interestingly, 10 of 11 regional banks belong to this class, which shows that they have a possibility to improve their efficiency by scaling up their activities. This observation leads us to hypothetically study the desirability of merging low ranked regional banks.

5.9.3 The Effects of a Merger

Confronted with the worst banking crisis in recent years, many banks have been forced to reconstruct their organizations and some have been nationalized under the control of the Japanese Financial Supervisory Agency. Mergers have always been thought of as a powerful tool for restructuring in competitive survival situations. In view of the above returns-to-scale characteristics of the inefficient regional banks, we will hypothetically merge two inefficient banks into one and evaluate changes in efficiency. The target banks are R9 and R10, which both belong to the same prefecture and have low CCR ranks (17th and 19th among 20). Their optimal weights and slacks for inputs, along with projections onto the CCR and the BCC efficient frontiers are listed in Table 5.4. Both banks have no slacks in output (operating net profits) and their projections remain at the status quo, so these values are not listed in the table. It is observed that both banks have positive weights in "Assets" (v_3^*), so a reduction in assets will contribute directly to improve their CCR – global – efficiencies. R10 has zero weights on "Branch" and "Employee," meaning that it needs to reduce these slacks in branch (15) and employee (206) before these inputs come to have positive weights and hence contribute to improving efficiencies. (Refer to Problem 3.5 for discussions on these subjects.) The CCR projection requires drastic reductions in inputs, especially for R10, while the BCC projection is moderate compared with the CCR case, especially for R10.

First, we simulate the case in which the BCC projected version of R9 and R10 merge to form a new bank — called R12. Hence R12 has 125 (=69+56) branches, 1,887 (=1,095+792) employees and assets of 1,419,675 (=1,215,757+600,796) yen to produce operating net profits 13,518 (= 9,993 + 3,525) yen. See Table 5.2. Although reduction of branches and employees is a traditional way of restructuring, there may be problems in reducing assets of banks. However, we think of this merger as providing a means of strengthening both banks by eliminating bad loans. In any case, we added this new bank to

Table 5.4. Weights, Slacks and Projections

Bank	Branch	CCR				Slacks Employee s_2^{-*}	Assets s_3^{-*}		
		Weights		Assets v_3^*	Branch s_1^{-*}				
		Employee v_1^*	Employee v_2^*						
R9	2.38E-3		0	5.02E-7	0	110	0		
R10		0	0	1.53E-7	15	206	0		

	CCR Projection			BCC Projection		
	Branch	Employee	Assets	Branch	Employee	Assets
R9	60 (-29%)	978 (-39%)	1,056,915 (-32%)	69 (-22%)	1095 (-32%)	1,215,757 (-22%)
R10	23 (-66%)	368 (-64%)	362,760 (-45%)	56 (-19%)	792 (-24%)	600,796 (-8%)

the data set and evaluated its efficiencies. The results are displayed in Table 5.5 along with those of the projected values of R9 and R10 denoted by R9' and R10', respectively.

Table 5.5. Efficiency of Projected and Merged Banks

Bank	Branch	Employee	Assets	CCR	BCC	RTS	Scale
R9'	69	1,095	1,215,757	0.877	1	Incr	0.877
R10'	56	792	600,796	0.604	1	Incr	0.604
R12	125	1,887	1,419,675	0.766	0.780	Incr	0.982

As expected, R9' and R10' come to full BCC efficiency and also improve their global CCR efficiencies. However, the merged bank R12 has a BCC efficiency, 0.780, which is less than 1, and is identified as still having increasing returns-to-scale status. This result for R12 leads us to conclude that merging two locally (BCC) efficient banks which have increasing returns-to-scale status results in a locally inefficient one, if the merged bank exhibits increasing returns-to-scale characteristics. Thus, we have,

Proposition 5.1 (Tone (1999)) *When two locally (BCC) efficient DMUs merge to form a new DMU having the sum of inputs and outputs of the two DMUs as its inputs and outputs, the new DMU is neither locally (BCC) nor globally (CCR) efficient, if increasing returns-to-scale prevails at all three DMUs.*

We do not prove this proposition here. Instead, we illustrate it in the case of a single input and single output using Figure 5.8, where R9', R10', R12 and R2 are symbolically plotted in order to help understand the situations. In the

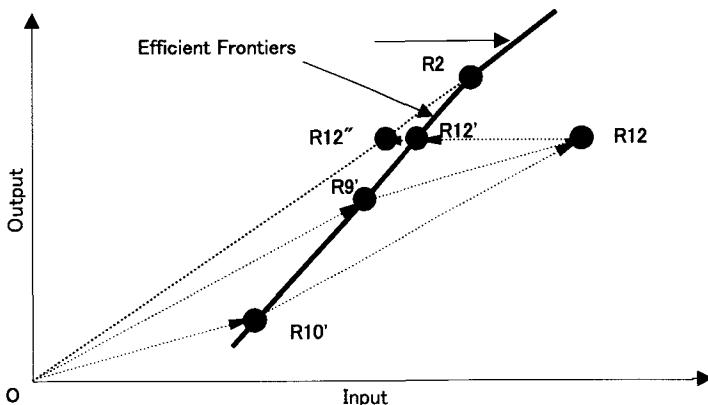


Figure 5.8. Merger is Not Necessarily Efficient.

figure, we merged R9' and R10' which when summed as a vector resulted in R12. This is a point inside the production possibility set of the BCC model and hence is locally inefficient. A further reduction in input is required to attain the BCC frontier point R12'.

This suggests that merging *and* restructuring may not be enough. Reduction of input resources is also necessary in order to arrive at a more efficient status. Otherwise, the merger will result in a worsened efficiency status.

To realize a globally efficient status is more difficult, as indicated by R12'' or R2 in the figure, the latter being attained by using the MPSS formula, as in (5.25).

As a second test, we used the data in Table 5.4 and merged the two banks R9 and R10 by projecting them onto the CCR frontiers. Comparing this result with R9 and R10 in Table 5.3 we find that this efficiency estimations result in a really drastic input reduction. The new bank has 83 ($=60+23$) branches, 1,346 ($=978+368$) employees, 1,419,675 ($=1,056,915+362,760$) assets to produce operating net profits 13,518 ($=9,993+3,525$). After adding this new bank to the data set, we evaluated its efficiency and found that the new bank attains global full efficiency.

In conclusion to this very limited study, we might say that inefficient regional banks are in urgent need of reconstruction not only by merging but also by reducing input resources, as well as by directing their marketing activities to new areas which city banks may not be able to approach. R2 is an instructive example. Otherwise, regional banks will not survive in competition against other financial institutions under present competitive conditions, which include the challenge of "global" worldwide efficiency evaluations.

5.10 SUMMARY

We have now covered the topics of returns to scale in its technical and mix efficiency aspects by reference to the MPSS concept. Scale inefficiencies were identified with failure to achieve MPSS. In the process we have given implementable form to the basic concepts, as drawn from the literature of economics, and extended them for simultaneous treatment of multiple outputs and inputs. This is needed for many (if not most) managerial uses. We have also eliminated the need for assumptions associated with achievement of technical efficiency, which are not likely to hold in practice. Finally, we provided interpretations and examples of their use which are intended to facilitate managerial, engineering, and social policy uses as in our example of bank merger problems and possibilities. We next extend this to other types of models and then ???? for uses in the next few sections. Later in this text we will further extend our treatments of returns to scale for use when unit prices and costs are available.

5.11 ADDITIVE MODELS

The model (5.22) which we used for MPSS, avoids the problem of choosing between input and output orientations, but this is not the only type of model for which this is true. The additive models to be examined in this section also have this property. That is, these models simultaneously maximize outputs and minimize inputs, in the sense of vector optimizations. The additive model we select is

$$\begin{aligned}
 & \max_{\lambda, s^-, s^+} \sum_{i=1}^m g_i^- s_i^- + \sum_{r=1}^s g_r^+ s_r^+ \\
 & \text{subject to} \\
 & \sum_{j=1}^n x_{ij} \lambda_j + s_i^- = x_{io}, \quad i = 1, 2, \dots, m \\
 & \sum_{j=1}^n y_{rj} \lambda_j + s_r^+ = y_{ro}, \quad r = 1, 2, \dots, s \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j, s_i^-, s_r^+ \geq 0.
 \end{aligned} \tag{5.46}$$

This model utilizes the “goal vector” approach of Thrall (1996a)¹⁰ in which the slacks in the objective are accorded “goal weights” which may be subjective or objective in character. Here we want to use these “goal weights” to ensure that the units of measure associated with the slack variables do not affect the optimal solution choices.

Employing the language of “dimensional analysis,” as in Thrall (1996a), we want these weights to be “contragradient” in order to insure that the resulting objective will be “dimensionless.” That is, we want the solutions to be free of

the units in which the inputs and outputs are stated. An example is the use of the input and output ranges in Cooper, Park and Pastor (1999)¹¹ to obtain $g_i = 1/R_i^-$, $g_r = 1/R_r^+$, where R_i^- is the range for the i^{th} input and R_r^+ is the range for the r^{th} output. This gives each term in the objective of (5.46) a contragradient weight. The resulting value of the objective is dimensionless, as follows from the fact that the s_i^- and s_r^+ in the numerators are measured in the same units as the R_i^- and R_r^+ in the denominators. Hence the units of measure cancel.

The condition for efficiency given in Chapter 1 for the CCR model is now replaced by the following simpler condition,

Definition 5.4 A DMU_o evaluated by (5.46) is efficient if and only if all slacks are zero.

To start our returns-to-scale analyses for these additive models we first replace the projections in (5.15) with

$$\begin{aligned}\hat{x}_{io} &= x_{io} - s_i^{-*} \quad i = 1, \dots, m \\ \hat{y}_{ro} &= y_{ro} - s_r^{+*}, \quad r = 1, \dots, s\end{aligned}\tag{5.47}$$

where s_i^{-*} and s_r^{+*} are optimal slacks obtained from (5.46). Then we turn to the dual (multiplier) model associated with (5.46) which we write as follows,

$$\begin{aligned}&\min_{v, u, u_0} \sum_{i=1}^m v_i x_{io} - \sum_{r=1}^s u_r y_{ro} \\ &\text{subject to} \\ &\sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} + u_0 \geq 0, \quad j = 1, \dots, n \\ &v_i \geq g_i^-, \quad u_r \geq g_r^+, \quad u_0 \text{ free},\end{aligned}\tag{5.48}$$

where, as was true for the BCC model, we will use variable u_0 to evaluate returns to scale.

We are thus in position to use Theorem 5.2 for “additive” as well “radial measures” as reflected in the BCC models discussed in earlier parts of this chapter. Hence we again have recourse to this theorem where, however, we note the difference in objectives between (5.11) and (5.46), including the change from $-u_0$ to $+u_0$. As a consequence of these differences we also modify (5.16) to the following,

$$\begin{aligned}&\max_{v, u, \hat{u}_0} \hat{u}_0 \\ &\text{subject to} \\ &\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - \hat{u}_0 \leq 0, \quad j = 1, \dots, n; \quad j \neq o \\ &\sum_{r=1}^s u_r \hat{y}_{ro} - \sum_{i=1}^m v_i \hat{x}_{io} - \hat{u}_0 = 0\end{aligned}\tag{5.49}$$

$$u_r \geq g_r^+, \quad v_i \geq g_i^-, \quad \hat{u}_0 \leq 0.$$

Here we have assumed that $u_0^* < 0$ was achieved in a first-stage use of (5.48). Hence, if $\hat{u}_0^* < 0$ is maximal in (5.49) then returns to scale are increasing at (\hat{x}_o, \hat{y}_o) in accordance with (i) in Theorem 5.2 whereas if $\hat{u}_0^* = 0$ then (iii) applies and returns to scale are constant at this point (\hat{x}_o, \hat{y}_o) on the efficient frontier.

For $\hat{u}_0^* > 0$ in stage one, the objective and the constraint on \hat{u}_0 are simply reoriented in the manner we now illustrate by using (5.46) to evaluate E in Figure 5.3 via

$$\begin{aligned} & \max_{\lambda, s^-, s^+} s^- + s^+ \\ & \text{subject to} \\ & \lambda_A + \frac{3}{2}\lambda_B + 3\lambda_C + 4\lambda_D + 4\lambda_E + s^- = 4 \\ & \lambda_A + 2\lambda_B + 4\lambda_C + 5\lambda_D + \frac{9}{2}\lambda_E - s^+ = \frac{9}{2} \\ & \lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_E = 1 \\ & s^-, s^+, \lambda_A, \lambda_B, \lambda_C, \lambda_D, \lambda_E \geq 0, \end{aligned}$$

where we have used unit weights for the g_i^- , g_r^+ , to obtain the usual additive model formulation. (See Thrall (1996b)¹² for a discussion of the applicable condition for a choice of such “unity” weights.) This has an optimal solution with $\lambda_C = \lambda_D = s^{-*} = 1/2$ and all other variables zero. To check that this is optimal we turn to the corresponding dual (multiplier) form for the above envelopment model which is

$$\begin{aligned} & \min 4v - \frac{9}{2}u + u_0 \\ & \text{subject to} \\ & v - u + u_0 \geq 0 \\ & \frac{3}{2}v - 2u + u_0 \geq 0 \\ & 3v - 4u + u_0 \geq 0 \\ & 4v - 5u + u_0 \geq 0 \\ & 4v - \frac{9}{2}u + u_0 \geq 0 \\ & v, u \geq 1, \quad u_0 : \text{free}. \end{aligned}$$

The solution $v^* = u^* = u_0^* = 1$ satisfies all constraints and gives $4v^* - \frac{9}{2}u^* + u_0^* = \frac{1}{2}$. This is the same value as in the preceding problem so that, by the duality theorem of linear programming, both solutions are optimal.

To determine the conditions for returns to scale we use (5.47) to project E into E' with coordinates $(\hat{x}, \hat{y}) = (\frac{7}{2}, \frac{9}{2})$ in Figure 5.3. Then we utilize the following reorientation of (5.49),

$$\min \hat{u}_0$$

subject to

$$\begin{aligned} v - u + \hat{u}_0 &\geq 0 \\ \frac{3}{2}v - 2u + \hat{u}_0 &\geq 0 \\ 3v - 4u + \hat{u}_0 &\geq 0 \\ 4v - 5u + \hat{u}_0 &\geq 0 \\ \frac{7}{2}v - \frac{9}{2}u + u_0 &\geq 0 \\ v, u &\geq 1, \quad \hat{u}_0 \geq 0. \end{aligned}$$

This also gives $v^* = u^* = \hat{u}_0^* = 1$ so the applicable condition is (ii) in Theorem 5.2. Thus returns to scale are decreasing at E', the point on the BCC efficient frontier which is shown in Figure 5.3.

5.12 MULTIPLICATIVE MODELS AND “EXACT” ELASTICITY

The treatments to this point have been confined to “qualitative” characterizations in the form of identifying whether RTS are “increasing,” “decreasing,” or “constant.” There is a literature — albeit a relatively small one — which is directed to “quantitative” estimates of RTS in DEA. Examples are Banker, Charnes and Cooper (1984),¹³ Førsund (1996)¹⁴ and Banker and Thrall (1992).¹⁵ However, there are problems in using the standard DEA models, as is done in these studies, to obtain scale elasticity estimates. Førsund (1996), for instance, lists a number of such problems. Also, the elasticity values in Banker and Thrall (1992) are determined only within upper and lower bounds. This is an inherent limitation that arises from the piecewise linear character of the frontiers for these models. Finally, attempts to extend the Färe, Grosskopf and Lovell (1985, 1994)¹⁶ approaches to the determination of “exact” scale elasticities have not been successful. See the criticisms in Førsund (1996, p.296) and Fukuyama (2000, p.105).¹⁷ (Multiple output-multiple input production and cost functions which meet the sub- and super-additivity requirements in economics are dealt with in Panzar and Willig (1977).¹⁸ See also Baumol, Panzar and Willig (1982).¹⁹

This does no, however, exhaust the possibilities. The class of “multiplicative models” opens additional opportunities. Introduced by this name in Charnes *et al.* (1982)²⁰ — see also Banker *et al.* (1981)²¹ — and extended in Charnes *et al.* (1983)²² To accord these models non-dimensional (=units invariance) properties. Although not used very much in applications these multiplicative models can provide advantages for extending the range of potential uses for DEA. For instance, they are not confined to efficiency frontiers which are concave. They can be formulated to allow the efficiency frontiers to be concave in some regions and non-concave elsewhere. See Banker and Maindiratta (1986).²³ They can also be used to obtain “exact” estimates of elasticities in manners that we now describe.

The models we use for this discussion are due to Banker and Maimdiratta (1986) — where analytical characterizations are supplied along with confirmation in controlled-experimentally designed simulation studies.

For this development we use an output oriented version of the multiplicative model which places us in consonance with the one in Banker and Maimdiratta (1986) — viz.,

$$\begin{aligned} & \max_{\gamma_0, \lambda} \gamma_0 \\ & \text{subject to} \\ & \prod_{j=1}^n x_{ij}^{\lambda_j} \leq x_{io} \quad i = 1, \dots, m \\ & \prod_{j=1}^n y_{rj}^{\lambda_j} \geq \gamma_0 y_{ro} \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \gamma_0, \lambda_j \geq 0, \end{aligned} \tag{5.50}$$

where “ \prod ” means “product of” and the variables with values to be determined are λ_j and γ_0 .

We convert these inequalities to equalities as follows.

$$\begin{aligned} e^{s_i^-} \prod_{j=1}^n x_{ij}^{\lambda_j} &= x_{io} \quad i = 1, \dots, m \\ \text{and} \\ e^{s_r^+} \prod_{j=1}^n y_{rj}^{\lambda_j} &= \gamma_0 y_{ro} \quad r = 1, \dots, s. \end{aligned} \tag{5.51}$$

We also replace the objective in (5.50) with $\gamma_0 e^{\varepsilon(\sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+)}$, where s_i^- , $s_r^+ \geq 0$ represent slacks. Employing (5.51) and taking logarithms convert this model to

$$\begin{aligned} & \min_{\lambda, \tilde{\gamma}_0, s^-, s^+} -\tilde{\gamma}_0 - \varepsilon \left(\sum_{r=1}^s s_r^+ + \sum_{i=1}^m s_i^- \right) \\ & \text{subject to} \\ & \tilde{x}_{io} = \sum_{j=1}^n \tilde{x}_{ij} \lambda_j + s_i^- \quad i = 1, \dots, m \\ & \tilde{\gamma}_0 + \tilde{y}_{ro} = \sum_{j=1}^n \tilde{y}_{rj} \lambda_j - s_r^+ \quad r = 1, \dots, s \\ & 1 = \sum_{j=1}^n \lambda_j \end{aligned} \tag{5.52}$$

$$\lambda_j, s_r^+, s_i^- \geq 0, \forall j, r, i,$$

where “~” denotes “logarithm” so the \tilde{x}_{ij} , \tilde{y}_{rj} , $\tilde{\gamma}_0$, \tilde{x}_{io} and \tilde{y}_{ro} are expressed in logarithmic units.

The dual to (5.52) is

$$\begin{aligned} & \max_{\alpha, \beta, \alpha_0} \sum_{r=1}^s \beta_r \tilde{y}_{ro} - \sum_{i=1}^m \alpha_i \tilde{x}_{io} - \alpha_0 \\ & \text{subject to} \\ & \sum_{r=1}^s \beta_r \tilde{y}_{rj} - \sum_{i=1}^m \alpha_i \tilde{x}_{ij} - \alpha_0 \leq 0, \quad j = 1, \dots, n \\ & \sum_{r=1}^s \beta_r = 1 \\ & \alpha_i \geq \varepsilon, \beta_r \geq \varepsilon; \quad \alpha_0 : \text{free}. \end{aligned} \tag{5.53}$$

Using α_i^* , β_r^* and α_0^* for optimal values, $\sum_{r=1}^s \beta_r^* \tilde{y}_{ro} - \sum_{i=1}^m \alpha_i^* \tilde{x}_{io} - \alpha_0^* = 0$ represents a supporting hyperplane (in logarithmic coordinates) for DMU_o, where efficiency is achieved. We may rewrite this log-linear supporting hyperplane in terms of the original input/output values:

$$\prod_{r=1}^s y_{ro}^{\beta_r^*} = e^{\alpha_0^*} \prod_{i=1}^m x_{io}^{\alpha_i^*}. \tag{5.54}$$

Then referring to (5.54) we introduce

Theorem 5.10 *Multiplicative Model RTS,*

- (i) *RTS are increasing if and only if $\sum_{i=1}^m \alpha_i^* > 1$ for all optimal solutions,*
- (ii) *RTS are decreasing if and only if $\sum_{i=1}^m \alpha_i^* < 1$ for all optimal solutions,*
- (iii) *RTS are constant if and only if $\sum_{i=1}^m \alpha_i^* = 1$ for some optimal solutions.*

To see what this means we revert to the discussion of (5.21) and introduce scalars a , b in (ax_o, by_o) . In conformation with (5.54) this means

$$e^{\alpha_0^*} \prod_{i=1}^m (ax_{io})^{\alpha_i^*} = \prod_{r=1}^s (by_{ro})^{\beta_r^*} \tag{5.55}$$

so that the thus altered inputs and outputs satisfy this extension of the usual Cobb-Douglas single output formulations.

The problem now becomes: given a specific expansion $a > 1$, contraction $a < 1$, or neither, i.e., $a = 1$, for application to all inputs, what is the value of b that positions the solution in the supporting hyperplane at this point? The answer is given by the following

Theorem 5.11 *If (ax_o, by_o) lies in the supporting hyperplane then $b = a \sum_{i=1}^m \alpha_i^*$.*

This theorem, as proved in Banker *et al.*(2003),²⁴ gives the precise result for the exact elasticity. Thus if $\sum_{i=1}^m \alpha_i^* = 1$ then $b = a$ and the outputs are expanded in the same proportion as the value of a which has been selected for the inputs. For $\sum_{i=1}^m \alpha_i^* > 1$ the proportional expansion in outputs exceed the proportion in which the inputs were expanded. Finally, for $\sum_{i=1}^m \alpha_i^* < 1$, the case of decreasing returns to scale applies since the outputs expansion is less than proportional to the proportion in which the inputs were expanded. Note that with any a specified for the inputs we then have a numerical value for the elasticity given by $\sum_{i=1}^m \alpha_i^*$.

There may be alternative optimal solutions for (5.53) so the values for the α_i^* components need not be unique. For dealing with alternate optima, we return to (5.50) and note that a necessary condition for efficiency is $\gamma_0^* = 1$. For full efficiency we must also have all slacks at zero in (5.51). Adaptation of (2.3), our previous approaches, projects to point on the efficient frontier via

$$\begin{aligned}\prod_{j=1}^n x_{ij}^{\lambda_j^*} &= e^{-s_i^*} x_{io} = x'_{io}, \quad i = 1, \dots, m \\ \prod_{j=1}^n y_{rj}^{\lambda_j^*} &= e^{s_r^+} y_{ro} = y'_{ro}, \quad r = 1, \dots, s\end{aligned}\tag{5.56}$$

and x'_{io} , y'_{ro} are the coordinates of the point on the efficiency frontier used to evaluate DMU_o.

Thus, we can extend the preceding models in a manner that is now familiar. Suppose we have obtained an optimal solution for (5.53) with $\sum_{i=1}^m \alpha_i^* < 1$. We then utilize (5.56) to form the following problem

$$\begin{aligned}\max_{\alpha, \beta, \alpha_0} \quad & \sum_{i=1}^m \alpha_i \\ \text{subject to} \quad & \sum_{r=1}^s \beta_r \tilde{y}_{rj} - \sum_{i=1}^m \alpha_i \tilde{x}_{ij} - \alpha_0 \leq 0, \quad j = 1, \dots, n; \quad j \neq o \\ & \sum_{r=1}^s \beta_r \tilde{y}_{ro} - \sum_{i=1}^m \alpha_i \tilde{x}_{io} - \alpha_0 = 0 \\ & \sum_{r=1}^s \beta_r = 1 \\ & \sum_{i=1}^m \alpha_i \leq 1 \\ & \alpha_i \geq \varepsilon, \quad \beta_r \geq \varepsilon; \quad \alpha_0 \text{ free in sign.}\end{aligned}\tag{5.57}$$

If $\sum_{i=1}^m \alpha_i^* = 1$ in (5.57), then returns to scale are constant by (iii) of Theorem 5.5. If the maximum is achieved with $\sum_{i=1}^m \alpha_i^* < 1$, however, condition (ii)

of Theorem 5.4 is applicable and returns to scale are decreasing at the point $x'_{io}, y'_{ro}; i = 1, \dots, m; r = 1, \dots, s$.

If we initially have $\sum_{i=1}^m \alpha_i^* > 1$ in (5.53), we replace $\sum_{i=1}^m \alpha_i \leq 1$ with $\sum_{i=1}^m \alpha_i \geq 1$ in (5.57) and also change the objective to minimize $\sum_{i=1}^m \alpha_i$. If the optimal value is greater than one, then (i) of Theorem 5.4 is applicable and RTS are increasing. On the other hand, if we attain $\sum_{i=1}^m \alpha_i^* = 1$ then condition (iii) applies and returns to scale are constant.

We can also derive pertinent scale elasticities in a straightforward manner. Thus, using the standard logarithmic derivative formulas for elasticities we obtain

$$\frac{d \ln b}{d \ln a} = \frac{a}{b} \frac{db}{da} = \sum_{i=1}^m \alpha_i^*. \quad (5.58)$$

Consisting of a sum of component elasticities, one for each input, this overall measure of elasticity is applicable to the value of the multiplicative expression with which DMU_o is associated.

The derivation in (5.58) holds only for points where this derivative exists. However, we can bypass this possible source of difficulty by noting that Theorem 5.10 allows us to obtain this elasticity estimate via

$$\frac{\ln b}{\ln a} = \sum_{i=1}^m \alpha_i^*. \quad (5.59)$$

Further, as discussed in Cooper, Thompson and Thrall (1996),²⁵ it is possible to extend these concepts to the case in which all of the components of y_o are allowed to increase by at least the factor b . That is, we need not assume that all outputs are scaled by the same values of a . However, we cannot similarly treat the constant, a , as providing an upper bound for the inputs since mix alterations are not permitted in the treatment of returns to scale in economics. See Varian (1984, p.20)²⁶ for RTS characterizations in economics.

In conclusion we turn to properties of units invariance for these multiplicative models. Thus we note that $\sum_{i=1}^m \alpha_i^*$ is units invariant by virtue of the relation expressed in (5.58). The property of units invariance is also exhibited in (5.59) since a and b are both dimension free. Finally, we also have

Theorem 5.12 *The models given in (5.50) and (5.51) are dimension free. That is, changes in the units used to express the input quantities x_{ij} or the output quantities y_{rj} in (5.50) will not affect the solution set or alter the value of $\max \gamma_0 = \gamma_0^*$.*

This theorem, which is proved in Banker *et al.* (2003), has the following corollary.

Corollary 5.3 *The restatement of (5.51) in logarithmic form yields a model which is translation invariant.*

5.13 SUMMARY OF CHAPTER 5

Although we have now covered all of the presently available models, we have not covered all of the orientations in each case. Except for the multiplicative models we have not covered output-oriented objectives for a variety of reasons. There are no real problems with the mathematical development but further attention must be devoted to how changes in input scale and input mix should be treated when all outputs are to be scaled up in the same proportions. See the discussion in Cooper, Thompson and Thrall (1996).

As also noted in Cooper, Thompson and Thrall (1996), the case of increasing returns to scale can be clarified by using Banker's most productive scale size to write $(\alpha \mathbf{x}_o, \beta \mathbf{y}_o)$. The case $1 < \beta/\alpha$ means that all outputs are increased by at least the factor β and returns to scale are increasing as long as this condition holds. The case $1 > \beta/\alpha$ has the opposite meaning — *viz.*, no output is increasing at a rate that exceeds the rate at which all inputs are increased. Only for constant returns to scale do we have $1 = \beta/\alpha$, in which case all outputs and all inputs are required to be increasing (or decreasing) at the same rate so no mix change is involved for the inputs.

The results in this chapter (as in the literature to date) are restricted to this class of cases. This leaves unattended a wide class of cases. One example involves the case where management interest is centered on only subsets of the outputs and inputs. A direct way to deal with this situation is to partition the inputs and outputs of interest and designate the conditions to be considered by $(\mathbf{x}_o^I, \mathbf{x}_o^N, \mathbf{y}_o^I, \mathbf{y}_o^N)$ where I designates the inputs and outputs that are of interest to management and N designates those which are not of interest (for such scale returns studies). Proceeding as described in the present chapter and treating \mathbf{x}_o^N and \mathbf{y}_o^N as "exogenously fixed," in the spirit of Banker and Morey (1986),²⁷ would make it possible to determine the situation for returns to scale with respect to the thus designated subsets. Also, as noted in Tone (2001),²⁸ the status of returns to scale may vary by the addition of the assurance region which will be introduced in the next chapter. Other cases involve treatments with unit costs and prices as in FGL (1994), Sueyoshi (1999),²⁹ and Tone and Sahoo.³⁰

The development covered in this chapter have been confined to technical aspects of production. Our discussions follow a long-standing tradition in economics which distinguishes scale from input mix changes by not allowing the latter to vary when scale changes are being considered. This permits the latter (i.e., scale changes) to be represented by a single scalar — hence the name. However, this can be far from actual practice, where scale and mix are likely to be varied simultaneously when determining the size and scope of an operation. See the discussion in Ray (2005)³¹ who uses the concept of "size" as distinct from "scale" as introduced in Maindiratta (1990)³² to study the efficiency of the banks in India.

There are, of course, many other aspects to be considered in treating returns to scale besides those attended to in the present chapter. Management efforts to maximize profits, even under conditions of certainty, require simulta-

neous determination of scale, scope and mix magnitudes with prices and costs known, as well as the achievement of the technical efficiency which is always to be achieved with *any* set of positive prices and costs. The topics treated in this chapter do not deal with such price-cost information. Moreover, the focus is on *ex-post facto* analysis of already effected decisions. This can have many uses, especially in the control aspects of management where evaluations of performance are required. Left unattended in this chapter, and in much of the DEA literature, is the *ex ante* (planning) problem of how to use this knowledge in order to determine how to blend scale and scope with mix and other efficiency considerations when efficiency future-oriented decisions. See Bogetoft and Wang (2005)³³ which extends DEA for use in evaluating proposed mergers in an *ex ante* fashion.

5.14 APPENDIX: FGL TREATMENT AND EXTENSIONS

In this Appendix, we first present the approach due to Färe, Grosskopf and Lovell (1985 and 1994). See also Färe and Grosskopf (1994). We then present a simple RTS approach without the need for checking the multiple optimal solutions as in Zhu and Shen (1995)³⁴ and Seiford and Zhu (1999)³⁵ where only the BCC and CCR models are involved. This approach will substantially reduce the computational burden, because it relies on the standard CCR and BCC computational codes (see Zhu (2002)³⁶ for a detailed discussion).

To start, we add to the BCC and CCR models by the following DEA model whose frontier exhibits non-increasing returns to scale (NIRS), as in Färe, Grosskopf and Lovell (FGL, 1985, 1994)

$$\begin{aligned} \theta_{NIRS}^* &= \min \theta_{NIRS} \\ \text{subject to} \\ \theta_{NIRS} x_{io} &= \sum_{j=1}^n x_{ij} \lambda_j + s_i^- \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ \quad r = 1, \dots, s \\ 1 &\geq \sum_{j=1}^n \lambda_j \\ 0 &\leq \lambda_j, \quad s_i^-, \quad s_r^+ \quad \forall i, r, j \end{aligned} \tag{5.60}$$

The FGL development rests on the following relation

$$\theta_{CCR}^* \leq \theta_{NIRS}^* \leq \theta_{BCC}^*$$

where “*” refers to an optimal value and θ_{NIRS}^* is defined in (5.60) while θ_{CCR}^* and θ_{BCC}^* refer to the BCC and CCR models.

FGL utilize this relation to form ratios that provide measures of RTS. However, we turn to the following tabulation which relates their RTS characterization to Theorems 2.3 and 2.4 (and accompanying discussion). See also Färe

and Grosskopf (1994), Banker, Chang and Cooper (1996), and Seiford and Zhu (1999)

	FGL Model	RTS	CCR Model
Case 1	If $\theta_{CCR}^* = \theta_{BCC}^*$	Constant	$\sum \lambda_j^* = 1$
Case 2	If $\theta_{CCR}^* < \theta_{BCC}^*$ then		
Case 3	If $\theta_{CCR}^* = \theta_{NIRS}^*$	Increasing	$\sum \lambda_j^* < 1$
Case 4	If $\theta_{CCR}^* < \theta_{NIRS}^*$	Decreasing	$\sum \lambda_j^* > 1$

It should be noted that the problem of non-uniqueness of results in the presence of alternative optima is not encountered in the FGL approach (unless output-oriented as well as input-oriented models are used) whereas they do need to be considered as in Theorem 5.4. However, Zhu and Shen (1995) and Seiford and Zhu (1999) develop an alternative approach that is not troubled by the possibility of such alternative optima.

We here present their results. See also Zhu (2002).

	Seiford and Zhu (1999)	RTS	CCR Model
Case 1	If $\theta_{CCR}^* = \theta_{BCC}^*$	Constant	$\sum \lambda_j^* = 1$
Case 2	If $\theta_{CCR}^* \neq \theta_{BCC}^*$ then		
Case 3	If $\sum \lambda_j < 1$ in any CCR outcome	Increasing	$\sum \lambda_j^* < 1$
Case 4	If $\sum \lambda_j > 1$ in any CCR outcome	Decreasing	$\sum \lambda_j^* > 1$

The significance of Seiford and Zhu's (1999) approach lies in the fact that the possible alternate optimal λ_j^* obtained from the CCR model only affect the estimation of RTS for those DMUs that truly exhibit constant returns to scale, and have nothing to do with the RTS estimation on those DMUs that truly exhibit increasing returns to scale or decreasing returns to scale. That is, if a DMU exhibits increasing returns to scale (or decreasing returns to scale), then $\sum \lambda_j^*$ must be less (or greater) than one, no matter whether there exist alternate optima of λ_j^* , because these DMUs do not lie in the MPSS region. This finding is also true for the u_0^* obtained from the BCC multiplier models.

Thus, in empirical applications, we can explore RTS in two steps. First, select all the DMUs that have the same CCR and BCC efficiency scores regardless of the value of $\sum_{j=1}^n \lambda_j^*$ obtained from model (2.5). These DMUs are constant returns to scale. Next, use the value of $\sum_{j=1}^n \lambda_j^*$ (in any CCR model outcome) to determine the RTS for the remaining DMUs. We observe that in this process we can safely ignore possible multiple optimal solutions of λ_j .

Part of the material in this chapter is adapted from Banker *et al.* (2004)³⁷ with permission from Elsevier Science.

5.15 RELATED DEA-SOLVER MODELS FOR CHAPTER 5

BCC-I, -O As explained in Chapter 4, these models solve the BCC model in the input-orientation and in the output-orientation, respectively. They also identify the returns-to-scale characteristics of DMUs which are recorded in the worksheet "RTS." For inefficient DMUs, we identify returns to scale

with the projected point on the efficient frontier. BCC-I uses the projection formula (4.17)-(4.18), while BCC-O uses the BCC version of the formula (3.74)-(3.75).

IRS-I, -O These codes solve the increasing (or non-decreasing) returns-to-scale model introduced in Section 5.7, i.e. the case $\sum_{j=1}^n \lambda_j \geq 1$ in input and output orientations.

DRS-I, -O These codes solve the decreasing (or non-increasing) returns-to-scale model in Section 5.7, i.e. the case $0 \leq \sum_{j=1}^n \lambda_j \leq 1$.

GRS-I, -O These correspond to the generalized returns-to-scale model, which has the constraint $L \leq \sum_{j=1}^n \lambda_j \leq U$. The values of $L(\leq 1)$ and $U(\geq 1)$ must be supplied through the Message-Box on the display by request.

5.16 PROBLEM SUPPLEMENT FOR CHAPTER 5

Problem 5.1

There is an ambiguity associated with characterizing B in Figure 5.3 as a point of constant returns-to-scale in that movement to the left produces a point on the efficient frontier where returns to scale is increasing. Can you supply a rationale for identifying B as a point of constant returns-to-scale?

Suggested Answer : Movement to the left from B would result in an output decrease that would be more than proportionate to the associated reduction in input so, generally speaking, one would need some justification for such a move. That is, the orientation would generally be toward increases in output so a movement which would reduce output more than proportionately to the associated input reduction would need justification. Similarly at C the orientation would favor movements to the left rather than the right.

Problem 5.2

Redraw Figure 5.3 to show what would happen to CCR- and BCC-efficiency if a new point, F , were added that is CCR-efficient. Discuss and analyze.

Suggested Answer : If the point F is on the broken line connecting B and C the MPSS region common to both BCC- and CCR-efficiency would be extended. An analytical characterization is as follows: Because F is CCR-efficient it would have an optimum with $\theta_o^* = 1$ and zero slacks for $\lambda_F^* = 1$ — the case of constant returns-to-scale in Theorem 5.3. This solution also satisfies the BCC model since $\sum_{j=1}^n \lambda_j^* = 1$. Because the CCR model is less constrained than the BCC model, we will *always* have optimal solutions with $\theta_o^* \leq \theta_B^*$. In this case, however, $\theta_o^* = \theta_B^*$. The solution must therefore be optimal for the BCC as well as the CCR model.

This is the crux of Theorem 5.5. As can be seen from Figure 5.3 and the numerical examples in Section 4, one can have $\theta_B^* = 1$ for the BCC model and $\theta_o^* < 1$ for the CCR model with zero slacks. This means that it is possible to have points which are BCC but not CCR efficient. As the theorem notes, it

is not possible to have the reverse situation so a point which is CCR-efficient must also be BCC-efficient.

Problem 5.3

Solve the hospital example in Table 1.5 in Chapter 1 using both (input-oriented) BCC-I and (output-oriented) BCC-O models in the DEA-Solver. Compare the returns-to-scale characteristics of hospitals that are BCC inefficient and are projected to the efficient frontiers by both models.

Suggested Answer : After computations you will find the following information by opening Worksheets “Score” and “RTS.”

(1) Input-oriented Case: The BCC-score, reference set and returns to scale (RTS) are exhibited in the left part of Table 5.6 below. Among the BCC efficient DMUs, A , B and D belong to CRS, while G , J , K and L to DRS. The returns-to-scale characteristics of BCC inefficient DMUs can be identified by referring to Theorem 5.8. For example, the reference set of DMU H consists of A , D and K of which A and D belong to CRS and K to DRS. Hence the projected activity of H displays DRS, since its reference set is a mixture of CRS and DRS. Similarly, DMUs F and I belong to DRS. However, the reference set of DMUs C and E belongs to CRS. So, we applied the procedure (i) on Page 133 to their projected activities and found that they belong to CRS.

Table 5.6. Results of Input-oriented/Output-oriented BCC Cases

DMU	Input-orientation			Output-orientation		
	Score	Ref Set	RTS	Score	Ref Set	RTS
A	1		CRS	1		CRS
B	1		CRS	1		CRS
C	0.8958	$B \ D$	CRS	0.9245	$B \ D \ G$	DRS
D	1		CRS	1		CRS
E	0.8818	$A \ B$	CRS	0.7666	$A \ B \ D \ L$	DRS
F	0.9389	$D \ J$	DRS	0.9548	$D \ J$	DRS
G	1		DRS	1		DRS
H	0.7988	$A \ D \ K$	DRS	0.8259	$A \ D \ K \ L$	DRS
I	0.9893	$A \ B \ L$	DRS	0.9899	$A \ B \ L$	DRS
J	1		DRS	1		DRS
K	1		DRS	1		DRS
L	1		DRS	1		DRS

(2) Output-oriented Case: The right half of Table 5.6 shows the results of the output-oriented BCC analysis. Since the output-oriented projection is different from the input-oriented case, BCC inefficient DMUs may have a

different reference set from the one in the input orientation. For example, DMU C 's reference set is composed of B , D and G in this case. Hence the output-oriented projection of C belongs to DRS, since the reference set is a mixture of CRS and DRS.

Notes

1. Throughout this chapter, we assume that the data set (X, Y) is semipositive.
2. R.D. Banker and R.M. Thrall (1992) "Estimating Most Productive Scale Size Using Data Envelopment Analysis," *European Journal of Operational Research* 62, pp.74-84.
3. R.D. Banker, I. Bardhan and W.W. Cooper (1996) "A Note on Returns to Scale in DEA," *European Journal of Operational Research* 88, pp.583-585.
4. R.D. Banker, H. Chang and W.W. Cooper (1996) "Equivalence and Implementation of Alternative Methods for Determining Returns to Scale in Data Envelopment Analysis," *European Journal of Operational Research* 89, pp.473-481.
5. R.D. Banker and R. Morey (1986) "Efficiency Analysis for Exogenously Fixed Inputs and Outputs" *Operations Research* 34, pp.513-521.
6. T. Ahn, A. Charnes and W.W. Cooper (1989), "A Note on the Efficiency Characterizations Obtained in Different DEA Models," *Socio-Economic Planning Sciences* 22, pp.253-257.
7. See R.D. Banker (1984), "Estimating Most Productive Scale Size Using Data Envelopment Analysis," *European Journal of Operational Research* 17, pp.35-44. See also R.D. Banker, A. Charnes and W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30, pp.1078-1092. A brief review of the history of this and related concepts is to be found in F.R. Førsund — "On the Calculation of the Scale Elasticity in DEA Models," *Journal of Productivity Analysis* 7, 1996, pp.283-302 — who traces the idea of a technically optimal scale to R. Frisch's classical volume, *Theory of Production* (Dordrecht: D.Reidel, 1965). See also L.M. Seiford and J. Zhu (1999) "An Investigation of Returns to Scale in Data Envelopment Analysis," *Omega* 27, pp.1-11.
8. This transformation is done explicitly in the Appendix to W.W. Cooper, R.G. Thompson and R.M. Thrall (1996) "Extensions and New Developments in DEA," *Annals of Operations Research* 66, pp.3-45.
9. K. Tone (1996) "A Simple Characterization of Returns to Scale in DEA" *Journal of the Operations Research Society of Japan* 39, pp.604-613.
10. R.M. Thrall (1996a), "Duality, Classification and Slacks in DEA," *Annals of Operations Research* 66, pp.109-138.
11. W.W. Cooper, K.S. Park and J.T. Pastor (1999), "RAM: A Range Adjusted Measure of Efficiency," *Journal of Productivity Analysis* 11, pp.5-42.
12. R.M. Thrall (1996b), "The Lack of Invariance of Optimal Dual Solutions under Translation Invariance," *Annals of Operations Research* 66, pp.109-138.
13. R.D. Banker, A. Charnes and W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Efficiencies in Data Envelopment Analysis," *Management Science* 30, pp.1078-1092.
14. F.R. Førsund (1996), "On the Calculation of Scale Elasticities in DEA Models," *Journal of Productivity Analysis* 7, pp.283-302.
15. R.D. Banker and R.M. Thrall (1992), "Estimation of Returns to Scale Using Data Envelopment Analysis," *European Journal of Operational Research* 62, pp.74-84.
16. R. Färe, S. Grosskopf and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer Nijhoff Publishing Co., and *Production Frontiers*, Cambridge University Press.
17. H. Fukuyama (2000), "Returns to Scale and Scale Elasticity in Data Envelopment Analysis," *European Journal of Operational Research* 125, pp.93-112.

18. J.C. Panzar and R.D. Willig (1977), "Economies of Scale in Multi-Output Production," *Quarterly Journal of Economics* XLI, pp.481-493.
19. W.J. Baumol, J.C. Panzar and R.D. Willig (1982), *Contestable Markets*, New York: Harcourt Brace Jovanovich.
20. A. Charnes, W.W. Cooper, L.M. Seiford and J. Stutz (1982), "A Multiplicative Model for Efficiency Analysis," *Socio-Economic Planning Sciences* 16, pp.213-224.
21. R.D. Banker, A. Charnes, W.W. Cooper and A. Schinnar (1981), "A Bi-Extremal Principle for Frontier Estimation and Efficiency Evaluation," *Management Science* 27, pp.1370-1382.
22. A. Charnes, W.W. Cooper, L.M. Seiford and J. Stutz (1983), "Invariant Multiplicative Efficiency and Piecewise Cobb-Douglas Envelopments," *Operations Research Letters* 2, pp.101-103.
23. R.D. Banker and A. Maindiratta (1986), "Piecewise Loglinear Estimation of Efficient Production Surfaces," *Management Science* 32, pp.126-135.
24. R.D. Banker, W.W. Cooper, L.M. Seiford, R.M. Thrall and J. Zhu (2003), "Returns to Scale In Different DEA Models," *European Journal of Operational Research* (to appear).
25. W.W. Cooper, R.G. Thompson and R.M. Thrall (1996), "Extensions and New Developments in DEA, Appendix," *The Annals of Operations Research* 66, pp.3-45.
26. H. Varian (1984), *Microeconomic Analysis*, New York: W.W. Norton.
27. R.D. Banker and R. Morey (1986), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* 34, pp.513-521.
28. K. Tone (2001), "On Returns to Scale under Weight Restrictions in Data Envelopment Analysis," *Journal of Productivity Analysis* 16, pp.31-47.
29. T. Sueyoshi (1999), "DEA Duality on Returns to Scale (RTS) in Production and Cost Analyses: An Occurrence of Multiple Solutions and Differences between Production-based and Cost-based RTS Estimates," *Management Science* 45, 1593-1608.
30. K. Tone and B.K. Sahoo (2005), "Cost-Elasticity: A Re-Examination in DEA," *Annals of Operations Research* (forthcoming).
31. S. Ray (2005), "Are Some Indian Banks too Large? An Examination of Size Efficiency in Indian Banking," *Journal of Productivity Analysis* (forthcoming).
32. A. Maindiratta (1990), "Scale and Size Efficiencies of Decision Making Units in Data Envelopment Analysis," *Journal of Econometrics* 46, pp.39-56.
33. P. Bogetoft and D. Wang (2005), "Estimating the Potential Gains from Mergers," *Journal of Productivity Analysis* 23, pp.145-171.
34. J. Zhu and Z. Shen (1995), "A Discussion of Testing DMU's Returns to Scale," *European Journal of Operational Research* 81, pp.590-596.
35. L.M. Seiford and J. Zhu (1999), "An Investigation of Returns to Scale under Data Envelopment Analysis," *Omega* 27, pp.1-11.
36. J. Zhu (2002), *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*, Kluwer Academic Publishers, Boston.
37. R.D. Banker, W.W. Cooper, L.M. Seiford, R.M. Thrall and J. Zhu, " Returns to scale in different DEA models," *European Journal of Operational Research* (2004), 154, pp.345-362.

6 MODELS WITH RESTRICTED MULTIPLIERS

6.1 INTRODUCTION

We have been dealing with several DEA models corresponding to different production possibility sets, which originate from technical aspects of organizational activities. In this manner we have minimized the need for “*a priori*” knowledge or the need for recourse to assumptions which are “outside the data.” There are situations, however, where additional information is available or where one is willing to make assumptions that lead to the imposition of conditions other than nonnegativity on the components of the multiplier vectors v and u . The beginning two sections of this chapter will treat these topics in terms of “assurance region” and “cone-ratio” approaches. In particular we will show how these assurance region and cone-ratio approaches can be incorporated in DEA models and how the results can be interpreted. This will be followed by applications designed to illustrate uses of these concepts.

For a preview we might note that Thompson, Singleton, Thrall and Smith (1986) developed the “assurance region” approach to help in choosing a “best site” for the location of a high-energy physics laboratory when other approaches proved to be deficient in evaluating outputs like “contributions to fundamental knowledge.”¹ Charnes, Cooper, Huang and Sun (1990)² developed the approach that they called “cone-ratio envelopment” in order to evaluate bank performances when unknown allowances for risk and similar factors needed to be taken into account.

6.2 ASSURANCE REGION METHOD

In the optimal weight (v_i^* , u_j^*) of DEA models for inefficient DMUs, we may see many zeros — showing that the DMU has a weakness in the corresponding items compared with other (efficient) DMUs. See Definition 2.1 and Problem 3.3 for troubles that can be associated with these efficiency characterizations. Large differences in weights from item to item may also be a concern. It was concerns like these that led to the development of the assurance region approach which imposes constraints on the relative magnitude of the weights for special items. For example, we may add a constraint on the ratio of weights for Input 1 and Input 2 as follows:

$$L_{1,2} \leq \frac{v_2}{v_1} \leq U_{1,2}, \quad (6.1)$$

where $L_{1,2}$ and $U_{1,2}$ are lower and upper bounds that the ratio v_2/v_1 may assume. The name *assurance region* (AR) comes from this constraint, which limits the region of weights to some special area. Generally, the DEA efficiency score in the corresponding envelopment model is worsened by additions of these constraints and a DMU previously characterized as efficient may subsequently be found to be inefficient after such constraints have been imposed.

We might note that the ratio of the multipliers is likely to coincide with the upper or lower bound in an optimal solution. Hence some care needs to be exercised in choosing these bounds so recourse to auxiliary information such as prices, unit costs, etc, is often used. Indeed, as we will later see, this approach provides a generalization of “allocative” and “price” efficiency approaches which require exact knowledge of prices and costs.

6.2.1 Formula for the Assurance Region Method

The assurance region, method is formulated for a DEA model by adding constraints like (6.1) for pairs of items if needed. For example, in the CCR model, the constraints (2.8)–(2.11) in Chapter 2 are augmented by the linear inequalities,

$$L_{1,2} v_1 \leq v_2 \leq U_{1,2} v_1. \quad (6.2)$$

Some authors like Roll, Cook and Golany³ have treated the bounds as absolute — viz., $L_i \leq v_i \leq U_i$ — rather than as ratios like in (6.1). This usage is readily accommodated by using the concept of a “numeraire”⁴ defined as a unit of account in terms of which all commodities (and services) are stated.

More generally, we can think of an “input numeraire” and an “output numeraire.” We can then constrain all of the values of the input (output) variables in the following manner.

$$v_1 l_{1,i} \leq v_i \leq v_1 u_{1,i} \quad (i = 2, \dots, m) \quad (6.3)$$

$$u_1 L_{1,r} \leq u_r \leq u_1 U_{1,r} \quad (r = 2, \dots, s) \quad (6.4)$$

Much in the manner in which money serves as a unit of account we can set $v_1 = u_1 = 1$ to obtain $(m + s - 2)$ constraints in absolute form. If some are not required, we can delete them from the constraints.

Thus, the CCR-AR model is:

$$(AR_o) \quad \max_{\mathbf{v}, \mathbf{u}} \quad \mathbf{u} \mathbf{y}_o \quad (6.5)$$

$$\text{subject to} \quad \mathbf{v} \mathbf{x}_o = 1 \quad (6.6)$$

$$-\mathbf{v} \mathbf{X} + \mathbf{u} \mathbf{Y} \leq \mathbf{0} \quad (6.7)$$

$$\mathbf{v} \mathbf{P} \leq \mathbf{0} \quad (6.8)$$

$$\mathbf{u} \mathbf{Q} \leq \mathbf{0} \quad (6.9)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, \quad (6.10)$$

where

$$\mathbf{P} = \begin{pmatrix} l_{12} & -u_{12} & l_{13} & -u_{13} & \dots & \dots & \dots & \dots \\ -1 & 1 & 0 & 0 & \dots & \dots & \dots & \dots \\ 0 & 0 & -1 & 1 & \dots & \dots & \dots & \dots \\ \dots & \dots \\ \dots & \dots \end{pmatrix}$$

and

$$\mathbf{Q} = \begin{pmatrix} L_{12} & -U_{12} & L_{13} & -U_{13} & \dots & \dots & \dots & \dots \\ -1 & 1 & 0 & 0 & \dots & \dots & \dots & \dots \\ 0 & 0 & -1 & 1 & \dots & \dots & \dots & \dots \\ \dots & \dots \\ \dots & \dots \end{pmatrix}.$$

However, notice that choice of the numeraire (v_1) in (6.1) is arbitrary so we could use constraints such as

$$l_{1,2} v_1 \leq v_2 \leq u_{1,2} v_1 \quad \text{and} \quad l_{2,3} v_2 \leq v_3 \leq u_{2,3} v_2$$

as well, with resulting in modifications of the matrices P and/or Q .

The computations should generally be carried out on the dual side since the envelopment model is usually easier to solve and interpret.

$$(DAR_o) \quad \min_{\theta, \boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\tau}} \quad \theta \quad (6.11)$$

$$\text{subject to} \quad \theta \mathbf{x}_o - X \boldsymbol{\lambda} + P \boldsymbol{\pi} \geq \mathbf{0} \quad (6.12)$$

$$Y \boldsymbol{\lambda} + Q \boldsymbol{\tau} \geq \mathbf{y}_o \quad (6.13)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\pi} \geq \mathbf{0}, \boldsymbol{\tau} \geq \mathbf{0}. \quad (6.14)$$

We now turn to relations between (AR_o) and (DAR_o) , assuming that both problems have a finite positive optimum. (Refer to Problem 6.5.) Let an optimal (and max-slack⁵) solution of (DAR_o) be $(\theta^*, \boldsymbol{\lambda}^*, \boldsymbol{\pi}^*, \boldsymbol{\tau}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$, where slacks \mathbf{s}^{-*} and \mathbf{s}^{+*} are defined in vector-matrix notation as:

$$\mathbf{s}^{-*} = \theta^* \mathbf{x}_o - X \boldsymbol{\lambda}^* + P \boldsymbol{\pi}^* \quad \text{and} \quad (6.15)$$

$$\mathbf{s}^{+*} = -\mathbf{y}_o + Y \boldsymbol{\lambda}^* + Q \boldsymbol{\tau}^*. \quad (6.16)$$

Based on the solution, we define “AR-efficiency” as:

Definition 6.1 (AR-Efficiency) *The DMU associated with (x_o, y_o) is AR-efficient, if and only if it satisfies*

$$\theta^* = 1, \quad s^{-*} = \mathbf{0} \quad \text{and} \quad s^{+*} = \mathbf{0}.$$

An improvement of an AR-inefficient (x_o, y_o) obtained by using (DAR_o) can be represented:

$$\hat{x}_o = \theta^* x_o - s^{-*} + P\pi^* \quad (= X\lambda^*) \quad (6.17)$$

$$\hat{y}_o = y_o + s^{+*} - Q\tau^* \quad (= Y\lambda^*). \quad (6.18)$$

These projection formulas now replace (3.28) and (3.29) to allow for the extra variables π^* and τ^* which appear in the envelopment model (DAR_o) as a result of the bounds placed on the variables in the multiplier model.

We now assert:

Theorem 6.1 *The activity (\hat{x}_o, \hat{y}_o) is AR-efficient.*

Proof. Since $(\hat{x}_o, \hat{y}_o) = (X\lambda^*, Y\lambda^*) \in P$ (the production possibility set), we can evaluate its AR-efficiency by solving the program:

$$(DAR_e) \quad \begin{aligned} & \min \quad \theta_e \\ & \text{subject to} \quad \theta_e \hat{x}_o - X\lambda_e + P\pi_e - s_e^- = \mathbf{0} \\ & \quad Y\lambda_e + Q\tau_e - s_e^+ = \hat{y}_o \\ & \quad \lambda_e \geq \mathbf{0}, \pi_e \geq \mathbf{0}, \tau_e \geq \mathbf{0}, s_e^- \geq \mathbf{0}, s_e^+ \geq \mathbf{0}. \end{aligned} \quad (6.19)$$

(DAR_e) has a feasible solution $(\theta_e = 1, \lambda_e = \lambda^*, \pi_e = \mathbf{0}, \tau_e = \mathbf{0}, s_e^- = \mathbf{0}, s_e^+ = \mathbf{0})$ so $0 \leq \theta_e^* \leq 1$. The dual to (DAR_e) is:

$$(AR_e) \quad \begin{aligned} & \max_{v_e, u_e} \quad u_e \hat{y}_o \\ & \text{subject to} \quad v_e \hat{x}_o = 1 \\ & \quad -v_e X + u_e Y \leq \mathbf{0} \\ & \quad v_e P \leq \mathbf{0} \\ & \quad u_e Q \leq \mathbf{0} \\ & \quad v_e \geq \mathbf{0}, \quad u_e \geq \mathbf{0}. \end{aligned} \quad (6.20)$$

Let $v_e^* = v^*/\theta^*$ and $u_e^* = u^*/\theta^*$, where (v^*, u^*) is an optimal solution of (AR_o) . Then, using these relations together with the complementary slackness conditions

$$v^* s^{-*} = v^* P\pi^* = u^* s^{+*} = u^* Q\tau^* = 0, \quad (6.21)$$

we obtain

$$v_e^* \hat{x}_o = \frac{v^*}{\theta^*} (\theta^* x_o - s^{-*} + P\pi^*) = v^* x_o = 1 \quad (6.22)$$

and

$$u_e^* \hat{y}_o = \frac{u^*}{\theta^*} (y_o + s^{+*} - Q\tau^*) = \frac{u^* y_o}{\theta^*} = 1. \quad (6.23)$$

Hence, the solutions ($\theta_e^* = 1, \lambda_e^* = \lambda^*, \pi_e^* = 0, \tau_e^* = 0, s_e^{-*} = 0, s_e^{+*} = 0$) and (v_e^*, u_e^*) are optimal to (DAR_e) and (AR_e) , respectively.

Moreover, we can demonstrate, in a similar way to the proof of Theorem 3.2, that every optimal solution of (DAR_e) satisfies $s_e^{-*} = 0$ and $s_e^{+*} = 0$. Thus, the activity (\hat{x}_o, \hat{y}_o) is AR-efficient. \square

We also need to note that the projection formulas (6.17, 6.18) do not always yield $\hat{x}_o \leq x_o$ and $\hat{y}_o \geq y_o$, contrary to (3.28) and (3.29). Also, the *output-oriented* assurance region model can be defined analogously to the output-oriented CCR model.

6.2.2 General Hospital Example

Tables 6.1 and 6.2 show data of 14 general hospitals with two inputs (Doctors and Nurses) and two outputs (Outpatients and Inpatients) and the optimal weights obtained by the CCR model where a number such as .332E-03 means $.332 \times 10^{-3}$. There are many zeros in the weights of inefficient DMUs, especially in the doctor and outpatient items. This means that the inefficient hospitals are very likely to have a surplus in doctors and shortage of outpatients, compared with the hospitals in the reference set.

Table 6.1. Data for 14 Hospitals

Hos.	Input		Output	
	Doctors	Nurses	Outpatients	Inpatients
H1	3008	20980	97775	101225
H2	3985	25643	135871	130580
H3	4324	26978	133655	168473
H4	3534	25361	46243	100407
H5	8836	40796	176661	215616
H6	5376	37562	182576	217615
H7	4982	33088	98880	167278
H8	4775	39122	136701	193393
H9	8046	42958	225138	256575
H10	8554	48955	257370	312877
H11	6147	45514	165274	227099
H12	8366	55140	203989	321623
H13	13479	68037	174270	341743
H14	21808	78302	322990	487539

To approach this problem by the assurance region method, we impose constraints

$$0.2 \leq \frac{v_2}{v_1} \leq 5, \quad 0.2 \leq \frac{u_2}{u_1} \leq 5. \quad (6.24)$$

Table 6.2. Efficiency and Weight of 14 Hospitals by CCR Model

Hospital	CCR Eff.	Doctor	Nurse	Outpatient	Inpatient
		v_1^*	v_2^*	u_1^*	u_2^*
H1	0.955	.332E-03	0	.959E-05	.167E-06
H2	1	.242E-03	.140E-05	.714E-05	.225E-06
H3	1	.104E-03	.204E-04	.339E-05	.325E-05
H4	0.702	.282E-03	.128E-06	0	.699E-05
H5	0.827	0	.245E-04	0	.384E-05
H6	1	.784E-04	.154E-04	.256E-05	.245E-05
H7	0.844	.133E-03	.102E-04	0	.505E-05
H8	1	.209E-03	.458E-07	.111E-07	.516E-05
H9	0.995	0	.233E-04	.426E-05	.136E-06
H10	1	.319E-04	.149E-04	.215E-05	.143E-05
H11	0.913	.162E-03	.739E-07	0	.402E-05
H12	0.969	.793E-04	.611E-05	0	.301E-05
H13	0.786	0	.147E-04	0	.230E-05
H14	0.974	0	.128E-04	0	.200E-05

The results are shown in Table 6.3.⁶ All the weights are now non zero, and in

Table 6.3. Efficiency and Weight of 14 Hospitals with Assurance Region Method

Hos.	AR Eff.	Doctor	Nurse	Outpa.		u_2^*/u_1^*	
				v_1^*	v_2^*	v_2^*/v_1^*	u_1^*
H1	0.926	.139E-03	.278E-04	0.2	.728E-05	.211E-05	0.29
H2	1	.110E-03	.219E-04	0.2	.575E-05	.167E-05	0.29
H3	1	.103E-03	.206E-04	0.2	.102E-05	.512E-05	5.0
H4	0.634	.116E-03	.232E-04	0.2	.116E-05	.579E-05	5.0
H5	0.820	.470E-05	.235E-04	5.0	.653E-06	.327E-05	5.0
H6	1	.776E-04	.155E-04	0.2	.280E-05	.224E-05	0.8
H7	0.803	.862E-04	.172E-04	0.2	.859E-06	.429E-05	5.0
H8	0.872	.794E-04	.159E-04	0.2	.790E-06	.395E-05	5.0
H9	0.982	.449E-05	.224E-04	5.0	.355E-05	.711E-06	5.0
H10	1	.519E-04	.114E-04	0.2	.284E-05	.863E-06	0.30
H11	0.849	.656E-04	.131E-04	0.2	.653E-06	.327E-05	5.0
H12	0.930	.516E-04	.103E-04	0.2	.513E-06	.257E-05	5.0
H13	0.740	.283E-05	.141E-04	5.0	.393E-06	.197E-05	5.0
H14	0.929	.242E-05	.121E-04	5.0	.336E-06	.168E-05	5.0

many instances the lower or upper bounds are achieved, as might be expected with linear (inequality) constraints. As might also be expected with extra variables in (6.12)-(6.14) CCR efficiency is generally reduced in value. For instance, H8 becomes inefficient in Table 6.3 whereas it was efficient in Table 6.2. This change in status is traceable to the change in weights resulting from the assurance region constraints. Thus H8 was efficient in Table 6.2 where it has comparatively large weights for doctors compared with nurses and a low weight to outpatients compared with inpatients. These weight unbalances allowed H8 to be efficient. But, the constraints of the assurance region prohibited such extreme weighting divergences and this resulted in a decline in efficiency. To portray the situation graphically Figure 6.1 provides a comparison of CCR models with and without the assurance region constraints.

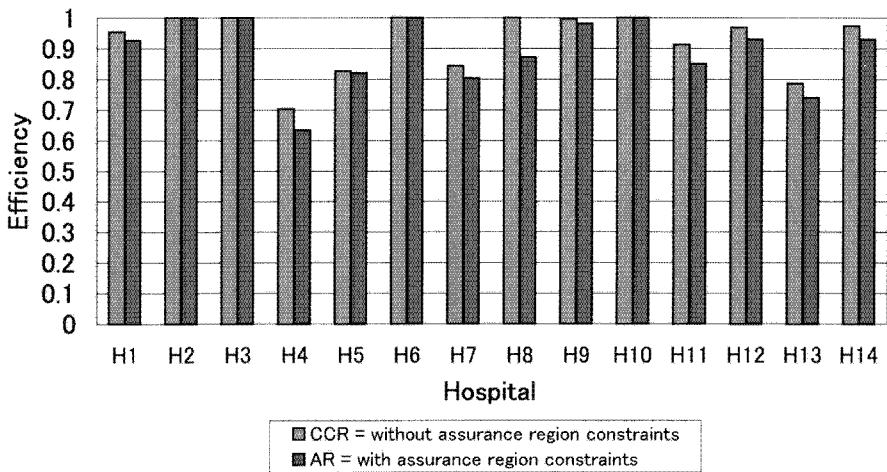


Figure 6.1. Efficiency Scores with and without Assurance Region Constraints

6.2.3 Change of Efficient Frontier by Assurance Region Method

To portray the influence of the assurance region method on the efficient frontier in the (v, u) -space we return to the case of Example 2.2 (two inputs and one output) represented by Figure 2.3 in Chapter 2. Now assume that we impose a constraint on the weights v_1 and v_2 ,

$$0.5 \leq \frac{v_2}{v_1} \leq 2. \quad (6.25)$$

That is, we require

$$0.5(v_1/u) \leq (v_2/u), \quad (v_2/u) \leq 2(v_1/u). \quad (6.26)$$

The feasible region in the (v, u) -space originally lay on or above the boundary indicated by the solid line in Figure 2.3 which changes to the area P in Figure

6.2 because of these constraints. The boundary of this new area P is composed of a part of line segments D and E , and two extremal rays corresponding to the constraints, and it is seen that the line segment $C : 8(v_1/u) + v_2/u = 1$ no longer contributes to P . In the original CCR model, C was efficient and, in fact, the values of v_1/u and v_2/u that make C efficient lay on the line segment $\overline{P_1 P_2}$ in Figure 2.3. Now, however, this line segment is shut out from the region P by the assurance region constraints. Hence, we move the line segment C in parallel to touch P (at P_5) and it is given by the dotted line

$$8(v_1/u) + (v_2/u) = 1/0.8. \quad (6.27)$$

Thus the efficiency of DMU C drops to 0.8. (Refer to DMU A in Figure 2.4.)

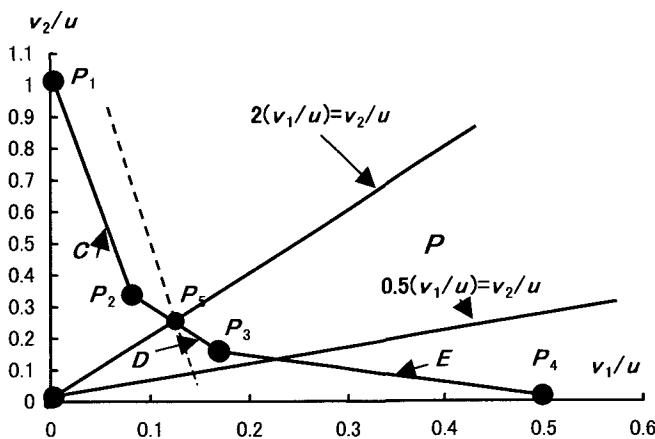


Figure 6.2. Assurance Region

6.2.4 On Determining the Lower and Upper Bounds

In this section we assume that all input and output items are technological (not monetary). Since, as was mentioned in the introductory chapter, the virtual inputs and outputs for the activity (x_k, y_k) are expressed as:

$$C_k = v_1 x_{1k} + \cdots + v_m x_{mk} \quad \text{and} \quad P_k = u_1 y_{1k} + \cdots + u_s y_{sk},$$

we can interpret v_i (u_j) as the unit cost (price) of the input (output) item x_{ik} (y_{jk}). This interpretation corresponds to the primal-dual LP relation of the variables $(\mathbf{v}$ vs. s^-) and $(\mathbf{u}$ vs. s^+) in which the optimal virtual multiplier v_i^* (u_j^*) is related to the imputed (shadow) price of the i (j)-th input (output) item. See the discussion of “complementary slackness” immediately following (AR_e) in (6.20), above.

Let c_{ik} be the actual unit cost of the input item x_{ik} . In view of this interpretation, it would be reasonable to set the lower and upper bounds for v_j/v_i

by the following scheme:

$$l_{ij} = \min \frac{c_{jk}}{c_{ik}} \quad (k = 1, \dots, n) \quad \text{and} \quad u_{ij} = \max \frac{c_{jk}}{c_{ik}} \quad (k = 1, \dots, n). \quad (6.28)$$

We can similarly put bounds on output multipliers by using the actual output prices. Returning to (6.1) we can now see more clearly some of the advantages of the assurance region approaches by virtue of the following considerations. In many cases, actual prices (or unit costs) are not known exactly. Nevertheless it may be possible to put bounds on their values as in (6.1). In other cases a variety of different levels of prices (or unit costs) may have occurred in which case bounds can be again used with the resulting solution of the AR problem selecting a value. If these values are not satisfactory new bounds may be used, and so on.

6.3 ANOTHER ASSURANCE REGION MODEL

As a generalization of the assurance region model, we introduce weight restrictions to inputs or outputs in a manner as described by the following example (see Allen *et al.*(1997)⁷) :

$$0.1 \leq \frac{v_1 x_1}{v_1 x_1 + v_2 x_2 + v_3 x_3} \leq 0.3 \quad (6.29)$$

$$0.05 \leq \frac{v_2 x_2}{v_1 x_1 + v_2 x_2 + v_3 x_3} \leq 0.1 \quad (6.30)$$

$$0.45 \leq \frac{v_3 x_3}{v_1 x_1 + v_2 x_2 + v_3 x_3} \leq 0.7 \quad (6.31)$$

$$0.4 \leq \frac{u_1 y_1}{u_1 y_1 + u_2 y_2} \leq 0.8 \quad (6.32)$$

In this example, we have three inputs x_1, x_2 and x_3 with respective weights v_1, v_2 and v_3 , and two outputs y_1 and y_2 with respective weights u_1 and u_2 . The ratio of the virtual weight of input 1 ($v_1 x_1$) against the total virtual weight ($v_1 x_1 + v_2 x_2 + v_3 x_3$) must be between the lower bound 0.1 and the upper bound 0.3. Similarly, $v_2 x_2 / (v_1 x_1 + v_2 x_2 + v_3 x_3)$ is restricted to the range (0.05, 0.1). We can impose the same weight restriction on outputs as well. In this example we have no explicit restriction imposed on $u_2 y_2$.

We call this model the *Assurance Region Global Model*.

Notice that (1) the sum of the lower bounds of inputs (outputs) must be less than or equal to unity and (2) the sum of the upper bounds must be greater than or equal to unity in the case all inputs (outputs) are restricted. We also recall the difference existing between this AR Global type model and the AR model. The (traditional) AR model imposes restriction on the virtual input multipliers, for instance,

$$L_2 \leq \frac{v_2}{v_1} \leq U_2 \quad (6.33)$$

$$L_3 \leq \frac{v_3}{v_1} \leq U_3. \quad (6.34)$$

This reduces to the following bounds.

$$\frac{x_1}{x_1 + U_2x_2 + U_3x_3} \leq \frac{v_1x_1}{v_1x_1 + v_2x_2 + v_3x_3} \leq \frac{x_1}{x_1 + L_2x_2 + L_3x_3}. \quad (6.35)$$

Thus, the relative weight of input 1 varies depending on the input values of the concerned DMU. The AR-Global model, however, restricts the relative weight to the designated (constant) bounds that are common to all DMUs. This scheme is intuitively more appealing than the traditional AR model.

6.4 CONE-RATIO METHOD

We now turn to the cone-ratio method which provides another approach. More general than the assurance region method, the cone-ratio approach lends itself to a variety of additional uses. However, we start by relating the cone-ratio to the assurance region approach.

6.4.1 Polyhedral Convex Cone as an Admissible Region of Weights

First we will observe the v -space. Let us restrict the feasible region of the weight v to be in the polyhedral convex cone spanned by the k admissible nonnegative direction vectors (\mathbf{a}_j) ($j = 1, \dots, k$). Thus, a feasible v can be expressed as

$$v = \sum_{j=1}^k \alpha_j \mathbf{a}_j \quad \text{with } \alpha_j \geq 0 \ (\forall j) \quad (6.36)$$

$$= A^T \boldsymbol{\alpha}, \quad (6.37)$$

where $A^T = (\mathbf{a}_1, \dots, \mathbf{a}_k) \in R^{m \times k}$ and $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_k)$. Let the polyhedral convex cone thus defined be V :

$$V = A^T \boldsymbol{\alpha}.$$

The two dimensional case is depicted in Figure 6.3.

Similarly, we restrict the feasible region of the output weight u in the polyhedral convex cone U spanned by the l admissible nonnegative direction vectors (\mathbf{b}_j) ($j = 1, \dots, l$) to define U ,

$$U = \sum_{j=1}^l \beta_j \mathbf{b}_j \quad \text{with } \beta_j \geq 0 \ (\forall j) \quad (6.38)$$

$$= B^T \boldsymbol{\beta}, \quad (6.39)$$

where $B^T = (\mathbf{b}_1, \dots, \mathbf{b}_l) \in R^{m \times l}$ and $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_l)$.

(Note 1.) We can demonstrate that the assurance region method is a special case of the cone-ratio method. In fact, if we impose a constraint such as

$$l_{1,2} \leq \frac{v_2}{v_1} \leq u_{1,2}$$

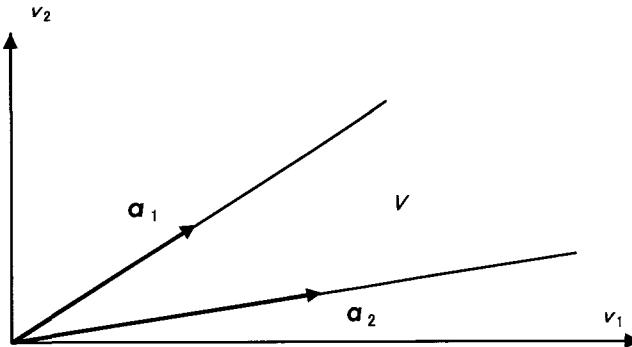


Figure 6.3. Convex Cone generated by Two Vectors

in the assurance region method, it corresponds to the selection of two admissible vectors in the cone-ratio method:

$$\mathbf{a}_1^T = (1, l_{1,2}, 0, 0, \dots, 0) \text{ and } \mathbf{a}_2^T = (1, u_{1,2}, 0, 0, \dots, 0),$$

where only two elements of the directional vectors are allowed to be positive. However, the cone-ratio method is more general in dealing with directions with all elements being possibly nonnegative.

6.4.2 Formula for Cone-Ratio Method

Given the polyhedral convex cones V and U generated by A and B , we have the following CCR model:

$$(CRP_o) \quad \max_{\mathbf{v}, \mathbf{u}} \quad \mathbf{u} \mathbf{y}_o \quad (6.40)$$

$$\text{subject to} \quad \mathbf{v} \mathbf{x}_o = 1 \quad (6.41)$$

$$-\mathbf{v} \mathbf{X} + \mathbf{u} \mathbf{Y} \leq \mathbf{0} \quad (6.42)$$

$$\mathbf{v} \in V \quad (6.43)$$

$$\mathbf{u} \in U, \quad (6.44)$$

with (CRP_o) coinciding with the ordinary CCR problem as represented in (3.2)-(3.7) when $V = R_m^+$ and $U = R_s^+$, the positive orthants in m and s space, respectively.

Using (6.37) and (6.39), we transform (CRP_o) into the following problem in row vectors α and β ,

$$(CRP'_o) \quad \max_{\alpha, \beta} \quad \beta(B \mathbf{y}_o) \quad (6.45)$$

$$\text{subject to} \quad \alpha(A \mathbf{x}_o) = 1 \quad (6.46)$$

$$-\alpha(AX) + \beta(BY) \leq \mathbf{0} \quad (6.47)$$

$$\alpha \geq \mathbf{0} \quad (6.48)$$

$$\beta \geq \mathbf{0}. \quad (6.49)$$

The dual of (CRD'_o) may be expressed in a real variable θ and a vector variable $\lambda = (\lambda_1, \dots, \lambda_n)^T$ as follows:

$$(CRD'_o) \quad \min_{\theta, \lambda} \quad \theta \quad (6.50)$$

$$\text{subject to} \quad \theta(Ax_o) - (AX)\lambda \geq 0 \quad (6.51)$$

$$(By_o) - (BY)\lambda \leq 0 \quad (6.52)$$

$$\lambda \geq 0. \quad (6.53)$$

From these expressions, it is clear that the polyhedral cone-ratio method also coincides with a CCR envelopment model evaluating the same DMUs but with the transformed positive data \bar{X} and \bar{Y} :

$$\bar{X} = AX \in R^{k \times n} \quad \text{and} \quad \bar{Y} = BY \in R^{l \times n}. \quad (6.54)$$

6.4.3 A Cone-Ratio Example

We will solve the same general hospital case as in Section 6.2.2. This has two input items (doctors and nurses) and two output items (outpatients and inpatients) covering 14 hospitals. The data set exhibited under (X^T, Y^T) in the first 4 columns of Table 6.4 has been transferred directly from Table 6.1.

Now let us choose two admissible input weight vectors, a_1 and a_2 , and one

Table 6.4. Efficiency of 14 Hospitals by CR (Cone-Ratio) and CCR Models

Hos.	Doctor	Nurse	Outpa.*	Inpa.**	Cone-Ratio			CR	CCR
	X^T		Y^T		\bar{X}^T		\bar{Y}^T	Eff.	Eff.
H1	3008	20980	97775	101225	13498	44968	401450	0.821	0.955
H2	3985	25643	135871	130580	16806.5	55271	527611	0.867	1
H3	4324	26978	133655	168473	17813	58280	639074	0.991	1
H4	3534	25361	46243	100407	16214.5	54256	347464	0.592	0.702
H5	8836	40796	176661	215616	29234	90428	823509	0.811	0.827
H6	5376	37562	182576	217615	24157	80500	835421	0.955	1
H7	4982	33088	98880	167278	21526	71158	600714	0.771	0.844
H8	4775	39122	136701	193393	24336	83019	716880	0.814	1
H9	8046	42958	225138	256575	29525	93962	994863	0.943	0.995
H10	8554	48955	257370	312877	33031.5	106464	1196001	1	1
H11	6147	45514	165274	227099	28904	97175	846571	0.809	0.913
H12	8366	55140	203989	321623	35936	118646	1168858	0.898	0.969
H13	13479	68037	174270	341743	47497.5	149553	1199499	0.714	0.786
H14	21808	78302	322990	487539	60959	178412	1785607	0.891	0.974

* Outpa.=Outpatients

** Inpa.=Inpatients

weight for output, \mathbf{b}_1 , as follows:

$$\begin{aligned}\mathbf{a}_1^T &= (1, 0.5) \\ \mathbf{a}_2^T &= (1, 2) \\ \mathbf{b}_1^T &= (1, 3).\end{aligned}$$

Thus, we have

$$A = \begin{pmatrix} 1 & 0.5 \\ 1 & 2 \end{pmatrix} \quad B = (1 \ 3).$$

The transformed data set ($\bar{X} = AX$ and $\bar{Y} = BY$) is also exhibited in Table 6.4 under the columns headed Cone-Ratio. We illustrate for H1,

$$\begin{aligned}\begin{pmatrix} 1 & 0.5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 3008 \\ 20980 \end{pmatrix} &= \begin{pmatrix} 13498 \\ 44968 \end{pmatrix} = \bar{X} \\ (1 \ 3) \begin{pmatrix} 97775 \\ 101225 \end{pmatrix} &= (401450) = \bar{Y}.\end{aligned}$$

Notice that the output data are compressed to one item.

By using these \bar{X} and \bar{Y} values, we can solve the cone-ratio model in its CCR form obtained by inserting \bar{X} and \bar{Y} in (6.51)-(6.53). The results are shown in the CR column in Table 6.4, along with the original CCR efficiency obtained from column 2 in Table 6.2. As can be seen only one hospital (H10) is cone-ratio efficient, while five hospitals were CCR efficient.

6.4.4 How to Choose Admissible Directions

There are several ways to choose (\mathbf{a}_j) and (\mathbf{b}_j) , from which the following pair are the most commonly used.

1. Use of the knowledge of experts for admissible ratios of weights.
2. First, solve the original CCR model and then select preferable DMUs among the efficient ones. Use the set of the optimal weights \mathbf{v}^* and \mathbf{u}^* for the preferable DMUs as the admissible directions.

In addition we can combine these two methods, as we now show, to obtain a third approach.

6.5 AN APPLICATION OF THE CONE-RATIO MODEL

We now provide an example of this third approach. The emphasis is on data adjustments as an alternative to the use of explicitly formulated constraints as in the assurance region approach. The objective is to deal with situations that are too complex to be dealt with by the latter method.

The problems we discuss are taken from Brockett *et al.* (1997),⁸ a study which is directed to providing an example for the Texas Banking Commission to consider for use in improving its monitoring activities. In particular it is

intended to provide an alternative to what is called the “risk adjusted capital” approach.

This concept, “risk adjusted capital,” stems from an agreement called the “Basel Agreement” for dealing with adjusting the net worth requirements of individual banks in order to allow for risk differences arising from the various components of each bank’s portfolio of loans and investments. This agreement was negotiated by the central banks of various countries with representatives meeting in Basel, Switzerland, in 1988. Subsequently adopted by both state and Federal government regulatory agencies in the U.S., this approach takes a form in which fixed percentages (the same for every bank) are applied to each class of items in a bank’s portfolio of loans and investments in order to obtain “jackup factors” to estimate the increase in capital (net worth) required to cover the risk associated with each class of items, such as loans, mortgages, stocks, bonds, etc.⁹ This is the origin of the term “risk adjusted capital.”

This is a very rigid approach. Hence the objective of the study we are examining was to explore DEA as a possible alternative (or supplement). However, the example is only illustrative and was designed to provide the Texas Banking Department with a “feel” for what DEA could do. Focusing on banks in Texas, the period 1984-1985 was selected for the data because it represented a period of transition. The years prior to 1984 were good ones for Texas banks, after which an increasing number of bank failures then began to appear. The cut-off in 1985 was used because this trend was only beginning to become apparent, at least mildly, in this year. See Table 6.5 and note that, by 1988, the increasing number of bank failures in Texas exceeded the number of failures in all of the other 49 states.¹⁰ The approach used by the Texas Banking Department failed to catch these developments, even when augmented by the risk adjusted capital approach, so this was considered to be a fair test for DEA.

Because this was an initiatory study, it was decided to confine the number of variables to the 4 inputs and the 4 outputs listed in Table 6.6. Reasons for the choice of these items are set forth in Brockett *et al.*(1997). We therefore here only break out some of the underlying components as shown in Table 6.6. Note that both balance sheet (stock) and income statement (flow) items are included whereas the Basel agreement is directed only to balance sheet items for their pertinence to risks and risk coverages. “Provision for Loan Losses” in Table 6.6 represents a flow item in the form of a charge against current income whereas “Allowance for Loan Losses” is a balance sheet item in the form of a “valuation reserve” which represents the cumulated amount of such “Provisions,” minus the loan losses charged to this reserve.¹¹

These input and output variables were selected in consultation with regulatory officials, and were intended to test the ability of DEA to allow for “risk” and “risk coverage” in its evaluations of efficient and inefficient performances. The inability of these officials to set bounds on specific variables made it impossible to use “assurance region” (or like) approaches. A strategy based on the cone-ratio envelopment approaches was therefore utilized which made it possible to use the expertise of these officials to identify a collection of “excellent

Table 6.5. Number of Bank Failures (through 10-31-88)

Year	Texas		U.S.	
	State banks	National banks	Total Texas	Nationwide
1980	0	0	0	10
1981	0	0	0	10
1982	4	3	7	42
1983	1	2	3	48
1984	2	4	6	79
1985	6	7	12	120
1986	14	12	26	145
1987	*24	28	52	188
1988	39	62	101	**175

* Includes 2 private uninsured banks.

** Includes 40 banks (including 9 State-chartered banks) which were closed, with some re-opening under a different aegis as a result of FDIC assistance.

Source: *Audit of the Examination and Enforcement Functions of Texas Department of Banking*, A report to the Legislative Audit Committee (Austin, Texas, Office of the State Auditor, January 1989).

banks" — which were not necessarily confined to banks in Texas. The idea was to use information derived from the performance of these excellent banks to reflect the risk present in their asset-liability (balance sheet) and operating (income statement) structures. It was believed that this could help to identify problems in Texas banks that fell short of excellence in their performances.

This objective was to be accomplished by using the multiplier values of these excellent banks in a cone-ratio envelopment to adjust the data of each of the individual banks included in this study. However, the following modification to this approach was also used. The "excellent banks" selected by the Banking Department representatives were submitted to DEA efficiency evaluation (using the CCR model) along with the other banks. Dual (multiplier) values for excellent banks found to be inefficient were not used to effect the desired data adjustments because it was deemed desirable to use only banks which were efficient as determined by DEA and excellent as determined by expert evaluations in order to reflect managerial performances as well as risks. See the discussion of "Effectiveness" vs. "Efficiency" following (3.96) in Chapter 3. See also R.M. Nun (1989),¹² a report which reflects a finding of the U.S. Comptroller of the Currency to the effect that bank management incompetence was a major factor in the failures that resulted in a crisis in U.S. banking in the late 1980s.

The five excellent banks as listed in Table 6.7, are (1) Bankers Trust Co. (2) Citibank N.A. (3) Morgan Guaranty Trust Co NY (4) Wachovia Bank and Trust Co. N.A. — as shown in rows 1-4 — and (5) First Interstate Bank of Nevada N.A., listed in row 21. However, 2 of these excellent banks were found

Table 6.6. Inputs and Outputs

INPUTS
Interest Expense
Interest expenses on deposit
Expense for federal funds purchased and repurchased in domestic offices
Non-Interest Expense
Salaries and employees benefits
Occupancy expense, furniture, and equipment
Provision for Loan Losses
Total Deposits
Sum of interest bearing and noninterest bearing deposits
OUTPUTS
Interest Income
Interest and fees on loans
Income on federal funds sold and repurchases in domestic offices
Total Non-Interest Income
Allowances for Loan Losses
Total Loans
Loans, net of unearned income

Source: Adjusted data from FDIC call reports are used in this study.

to be inefficient in each year. See the discussion of Table 6.7 below. Therefore the number of excellent banks used to supply multiplier values was reduced from 5 to the 3 following ones in each year. For 1984: 1 (=Bankers Trust), 4 (=Wachovia) and 21 (=First Interstate of Nevada). For 1985: 1 (=Bankers Trust), 3 (=Morgan guaranty) and 21 (=First Interstate of Nevada).

Thus for each year we have a matrix of the form

$$D = \begin{pmatrix} v_{11}^* & \dots & v_{41}^* & O \\ \dots & & & \\ v_{13}^* & \dots & v_{43}^* & u_{11}^* & \dots & u_{41}^* \\ O & & & \dots & & \\ & & u_{13}^* & \dots & u_{43}^* & \end{pmatrix} \quad (6.55)$$

with the stars designating averages of the optimal multipliers represented by u^* , to be entered as elements of the B matrix of (6.51)-(6.52), and the v^* to be entered as elements of the A matrix so that

$$D \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} AX \\ BY \end{pmatrix} = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix}, \quad (6.56)$$

as given in (6.54).

Table 6.7. CCR and Cone-Ratio Efficiency Scores (1984, 1985)*

No.	Bank title	1984		1985	
		CCR	Efficiency Scores Cone-R.	CCR	Efficiency Scores Cone-R.
1	BANKERS TRUST CO	1.0000	1.0000	1.0000	1.0000
2	CITIBANK NA	1.0000	1.0000	1.0000	0.9575
3	MORG. GUARANTY TR CO NY	0.9757	0.9985	1.0000	1.0000
4	WACHOVIA BK and TR CO NA	1.0000	1.0000	1.0000	0.9505
5	INTERFIRST BK AUSTIN NA	1.0000	0.9987	1.0000	0.7949
6	TX COMM BK AUSTIN NA	1.0000	0.9865	1.0000	0.9915
7	FIRST CITY BK OF DALLAS	1.0000	0.9997	1.0000	0.9947
8	INTERFIRST BK DALLAS NA	1.0000	0.9623	1.0000	0.9925
9	MBANK DALLAS NA	1.0000	0.9512	1.0000	0.8375
10	REPUBLIC BK DALLAS NA	0.9687	0.9289	1.0000	0.8260
11	INTERFIRST BK FT WOR. NA	1.0000	0.9650	0.9464	0.9017
12	TX AMER. BK FT WOR. NA	1.0000	1.0000	1.0000	0.9213
13	ALLIED BK OF TX	0.8373	0.8556	0.8158	0.6632
14	CAPITAL BK NA	1.0000	0.9943	1.0000	0.7864
15	FIR CITY NAT BK HOUSTON	1.0000	0.9632	1.0000	0.9812
16	INTERFIR BK HOUSTON NA	1.0000	0.9880	1.0000	0.9812
17	REPUBLIC BK HOUSTON NA	0.7794	0.7208	0.8585	0.7254
18	TX COMMERCE BK NA	1.0000	0.9260	1.0000	0.9034
19	FROST NA BK SAN ANTONIO	1.0000	0.9000	0.9568	0.7411
20	NB OF COMM SAN ANTONIO	0.9044	0.9513	0.9621	0.9794
21	FIRST INTRST BK NEVADA NA	1.0000	1.0000	1.0000	1.0000

* Source: P.L. Brockett, A. Charnes, W.W. Cooper and Z.M. Huang (1997). See Note 8 at the end this chapter.

Using these results the following (ordinary) CCR model was achieved for use on the thus transformed data

$$\begin{aligned}
 & \min \quad \theta \\
 \text{subject to} \quad & \theta \bar{x}_o \geq \bar{X} \lambda \\
 & \bar{y}_o \leq \bar{Y} \lambda \\
 & \lambda \geq 0.
 \end{aligned} \tag{6.57}$$

The results from the transformed data are shown in the second of the two columns exhibited in Table 6.7. The first of the two columns portrays results from a use of the CCR model on the untransformed (original) data so the comparisons can easily be made. As can be seen, some of the excellent banks have

their efficiency scores changed with Morgan Guaranty improving its efficiency score in 1985 while Citibank and Wachovia in 1985 leave the status of being fully efficient. More importantly, many of the Texas banks lose the status of full efficiency which was accorded to them in 1984 when the original data were transformed by the cone-ratio envelopment used for 1984. More importantly, all of the Texas banks lose their efficiency status in 1985. In addition not a single Texas bank improved its efficiency score — although improvement is possible with these transformed data as is seen in the case of Morgan Guaranty. Moreover, every one of the Texas banks worsened its score when going from 1984 to 1985. Taken together this suggests that a worsening banking situation began to develop in 1984, as judged by the excellent-efficient performances present in the transformed data and the situation deteriorated further in 1985.

6.6 NEGATIVE SLACK VALUES AND THEIR USES

Table 6.8, below, allows us to delve deeper into these results in order to better understand what is happening with these transformations. This table provides details underlying the 1985 evaluation of Interfirst Bank of Fort Worth — the bank exhibited on line 11 in Table 6.7 — in the following manner. Column 1 shows the actual (non-transformed) data for this bank and column 2 shows the performance that was required to achieve efficiency under the heading “Value if Efficient.”

As can be seen, all of the outputs were satisfactory but the only satisfactory performance among the 4 inputs appears for “deposits” in the final line. Both “interest expenses” and “non interest expenses” should have been reduced by the amounts shown in the column headed “Potential Improvement.”

Special interest attaches to the Provision for Loan Losses an input which is used to cover potential risks due to this bank’s lending activities. As can be seen, the value for this item is negative under “Potential Improvement.” Hence the amount of this *input* should have been *increased* by \$4.7 (million).

The work of this study was suspended when part of the study task force was invited by the Texas State Auditor’s Office to make a similar study for use in an audit of the State’s Insurance Department. Hence the future extensions needed to perfect this DEA model for use in banking were never undertaken. Here we only note that this could include added constraints connecting balance sheet items such as Allowances for Loan Losses to corresponding income statement items like “Provision for Loan Losses.” Nevertheless, the potential for future improvement is present so we study this topic a bit further as follows.

Consider the following reformulation of the constraints of (6.57),

$$\begin{aligned} s^+ &= (BY)\lambda - By_o && (6.58) \\ s^- &= -(AX)\lambda + \theta(Ax_o) \\ 0 &\leq \lambda, s^+, s^-. \end{aligned}$$

This is in the same form as (6.51)-(6.53) with the inclusion of nonnegativity for the slacks.

Table 6.8. Printout for Cone-Ratio CCR Model - Interstate Bank of Fort Worth, 1985.

DECISION MAKING UNIT: 11 1985 INTERFIRST OF FORT WORTH				
IN MILLION DOLLARS	ACTUAL	VALUE IF EFFICIENT	POTENTIAL IMPROVEMENT	SUM OF IMPROVEMENTS
OUTPUTS				
INTINCOME	111.61	111.61	0.00	
NONINTINC	20.38	20.38	0.00	
ALLOWANCE	16.30	16.30	0.00	
NETLOANS	970.43	970.43	0.00	0.00
INPUTS				
INTEXPENS	91.51	79.51	12.00	
NONINTEXP	26.61	11.49	15.12	
PROVISION	7.00	11.70	-4.70	
DEPOSITS	1227.41	1227.41	0.00	22.42
				\$22.42 MILLIONS

We now write the following version for the CCR constraints to be used on the original data as follows,

$$\begin{aligned}\tilde{s}^+ &= Y\lambda - y_o \\ \tilde{s}^- &= -X\lambda + \theta x_o,\end{aligned}\tag{6.59}$$

where $\mathbf{0} \leq \lambda$ but the slacks are not otherwise constrained.

To show why these nonnegativity requirements are not imposed on the slacks in (6.59), we simplify matters by assuming that the A and B matrices have inverses.

Applying these inverses, A^{-1} and B^{-1} , to (6.58) produces

$$\begin{aligned}B^{-1}s^+ &= Y\lambda - y_o \\ A^{-1}s^- &= -X\lambda + \theta x_o \\ \mathbf{0} \leq \lambda, s^+, s^- &\end{aligned}\tag{6.60}$$

Setting $\tilde{s}^+ = B^{-1}s^+$, $\tilde{s}^- = A^{-1}s^-$ we have a solution to (6.59). Moreover, if this solution is optimal for (6.58) then it is optimal for (6.59) since otherwise we could reverse this procedure and show that it is not optimal for (6.58).

Thus we have a solution of (6.60), in terms of original data, derived from (6.58) in terms of transformed data. However, we cannot guarantee that the slacks in (6.58) will satisfy nonnegativity. If this is wanted one needs to join the constraints in (6.58) and (6.59) to form the following augmented problem,

$$s^+ = (BY)\lambda - By_o\tag{6.61}$$

$$\begin{aligned}s^- &= -(AX)\lambda + \theta(Ax_o) \\ \tilde{s}^+ &= Y\lambda - y_o \\ \tilde{s}^- &= -X\lambda + \theta x_o \\ 0 \leq \lambda, s^+, s^-, \tilde{s}^+, \tilde{s}^- &\leq\end{aligned}$$

In this example we have assumed the existence of the ordinary inverse A^{-1} , B^{-1} . However, this could be extended to more general inverses¹³ to show how the above derivation, from transformed to original data, could be made without this assumption. Finally, it is also possible to modify the problem (6.61) so that nonnegativity is imposed on only some of the variables in \tilde{s}^+ , \tilde{s}^- . This, in fact, was what was done in this study and the slack variables for "Provision" was the only one not subject to a nonnegativity constraint. See Problem 6.2.

6.7 A SITE EVALUATION STUDY FOR RELOCATING JAPANESE GOVERNMENT AGENCIES OUT OF TOKYO

Takamura and Tone¹⁴ have developed a consensus-making method for a national project in Japan. This section outlines the project in a manner that puts emphasis on the methodological aspects.

6.7.1 *Background*

Tokyo, with its population of more than 10 million persons is one of the largest megalopolises in the world. It serves as the center of economics and politics in nearly monopoly fashion in Japan. However, the efficiency of its functions as a city, e.g., culture, residence, transportation, health care, and anti-natural disaster measures, have dropped dramatically during the last two decades. In an effort to solve such urban problems, the Diet of Japan has decided to devolve some governmental functions from Tokyo, and a bill to transfer the Diet, Governmental Agencies and the Supreme Court to a new capital was approved by the Diet in 1992. The aim of this separation is to create a Washington D.C. in Japan, and to allow Tokyo to maintain itself in a manner similar to New York. This is a huge long-range national project with an initial budget more than 12 trillion yen (about 10 billion U.S. dollars).

A council for supervising this transfer to a new capital has been organized which consists of 19 "wise men" appointed by the Prime Minister. By the end of 1998, the Council selected 10 candidate sites for the new capital. They are Miyagi (A), Fukushima (B), Tochigi (C), Tochigi/Fukushima (D), Ibaraki (E), Shizuoka/Aichi (F), Gifu/Aichi (G), Mie (H), Mie/Kio (I) and Kio (J). Under the Council, a committee for evaluating these sites was organized in 1998 consisting of 6 experts in decision making, economics, law, civil engineering, environmental issues and assessment. One of the objects of this Committee is to survey potential methods for consensus formation among the Council members. The following method was proposed by Tone, as a Committee member, for this purpose.

6.7.2 The Main Criteria and their Hierarchy Structure

The Council selected the following three categories of criteria as crucial for evaluating the sites for the location of the new capital:

Influence on the future of the country:

Subjects related to the long-range strategic aspects of the relocation as they affect the entire country.

Conditions for establishing the capital city functions:

Terms related to the capital city functions.

Conditions for suitability of the new location:

Terms related to the suitability and feasibility of the relocation.

These three factors are further divided into eighteen concrete and detailed criteria (C1-C18) via two intermediate criteria (B1 and B2) as follows:

1. Influence on the future of the country:

- (B1) Direction for reorganizing the national structure.
This criterion is divided into:
 - (C1) Direction of reform of national land structure.
 - (C2) Rectification of the excessive concentration of activities in Tokyo.
- (C3) Direction of culture formation.
- (C4) Ease of correspondence with a new information network.
- (C5) Speedy response in a large-scale disaster.

2. Conditions for establishing the capital city functions:

- (C6) Ease of access to foreign countries.
- (C7) Ease of access to Tokyo.
- (C8) Ease of access to the entire country.
- (C9) Appeal of the landscape.
- (C10) Safety in the event of an earthquake disaster.
- (C11) Safety in the event of a volcanic eruption.

3. Conditions for suitability of the new location:

- (C12) Feasibility of the smooth acquisition of land.
- (C13) Suitability of topographical features.
- (C14) Safety against flood and sediment disasters.
- (C15) Stability of water supply.
- (C16) Suitability in terms of relations with existing cities.

- (B2) Environmental issues

This criterion is divided into:

- (C17) Harmony with the natural environment.
- (C18) Possibility of lessening the environmental load.

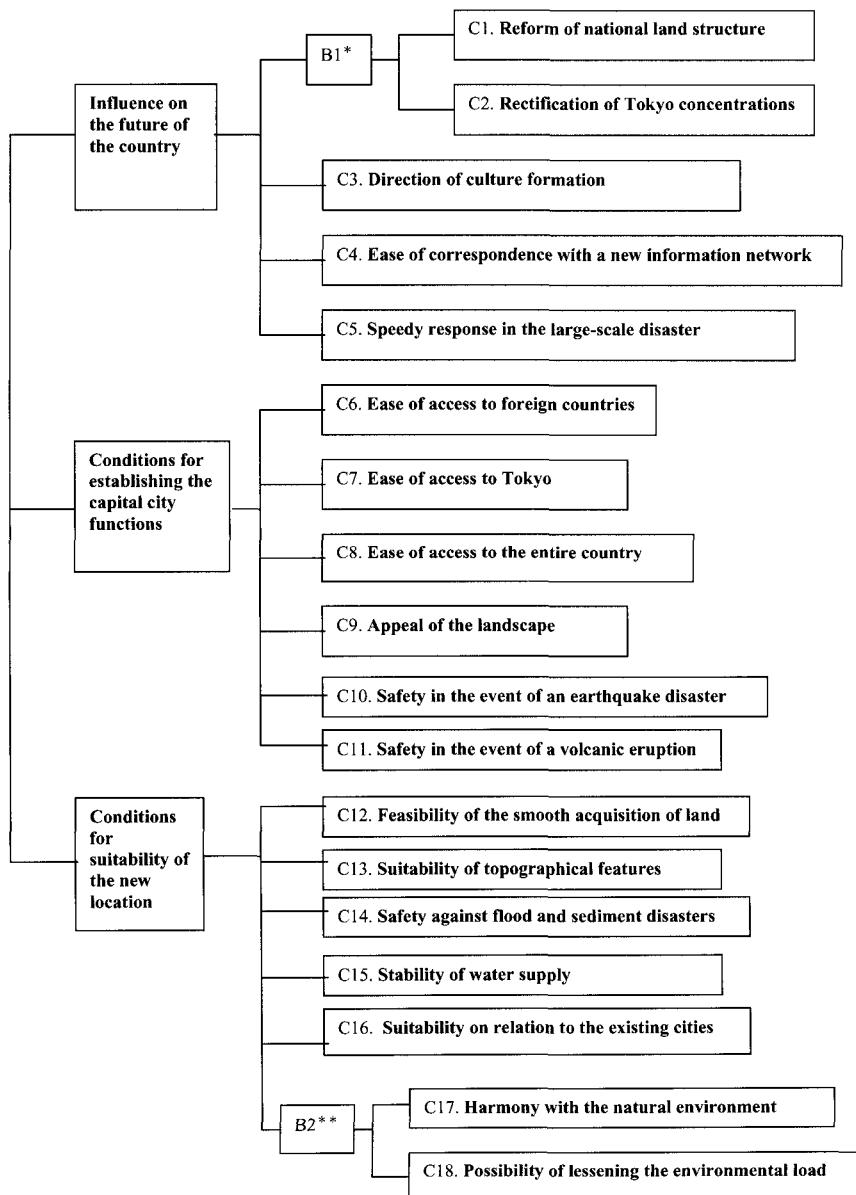
Figure 6.4 summarizes the hierarchical structure expressed above.

6.7.3 Scores of the 10 Sites with respect to the 18 Criteria

Table 6.9 exhibits the score matrix $S = (S_{ij})$ ($i = 1, \dots, 18 : j = 1, \dots, 10$) of the 10 candidate sites (A-J) with respect to the 18 criteria (C1-C18). These values were obtained from expert teams consisting of five specialists, on average. For example, (C4) (“Ease of correspondence with a new information system”) was investigated by six members headed by Professor T. Ishii of Keio University, (C6), (C7) and (C8) (“Access to other areas”) are evaluated by five members headed by Professor S. Morichi of Tokyo Institute of Technology. (C10) and (C11) (“Earthquake and volcano”) are reviewed by ten members headed by Professor M. Fuchigami of University of Tokyo. The scores were measured by an absolute value ranging from 5 (the best) to 1 (the worst).

Table 6.9. Scores (S_{ij}) of 10 Sites (A-J) with respect to 18 Criteria (C1-C18)

	A	B	C	D	E	F	G	H	I	J	Avg	SD
C1	2.8	2.6	3.0	2.8	2.4	2.8	2.9	2.4	2.4	2.4	2.7	0.23
C2	3.0	3.0	3.0	3.0	2.5	3.5	3.5	4.0	4.0	4.0	3.4	0.53
C3	2.9	2.9	3.4	3.3	2.4	2.8	3.0	2.4	2.3	2.4	2.8	0.38
C4	3.0	2.3	2.3	2.3	3.0	4.3	4.3	2.3	2.3	2.3	2.8	0.82
C5	4.0	4.0	4.0	4.0	4.0	2.0	3.5	3.0	3.0	3.0	3.5	0.69
C6	2.2	2.1	2.5	2.5	3.7	4.2	4.4	5.0	3.8	3.8	3.4	1.02
C7	3.3	3.5	5.0	5.0	4.2	3.5	2.3	2.1	2.0	2.0	3.3	1.18
C8	3.4	3.3	3.9	3.9	3.6	5.0	4.7	4.3	4.3	4.3	4.1	0.55
C9	2.5	2.5	5.0	5.0	2.0	4.0	3.0	3.5	3.0	3.0	3.4	1.03
C10	4.0	5.0	4.0	4.0	5.0	1.0	2.0	2.0	2.0	2.0	3.1	1.45
C11	2.6	3.2	1.3	2.6	4.4	4.2	4.7	5.0	5.0	5.0	3.8	1.30
C12	3.6	2.8	3.6	3.2	3.4	2.0	4.1	1.6	2.6	3.3	3.0	0.76
C13	2.3	3.0	4.7	4.0	4.1	3.2	2.3	4.0	3.3	3.1	3.4	0.78
C14	3.7	4.2	3.7	4.1	3.0	2.9	4.5	3.0	3.6	3.6	3.6	0.55
C15	3.5	2.5	2.5	2.5	1.5	3.0	3.0	3.5	2.2	1.5	2.6	0.71
C16	4.2	3.3	3.0	3.2	2.4	4.1	3.5	3.5	3.5	3.5	3.4	0.52
C17	2.9	3.3	3.6	4.1	2.7	2.5	1.9	2.4	2.4	2.1	2.8	0.68
C18	2.6	3.8	3.0	3.6	3.4	4.2	3.0	2.9	2.6	2.3	3.1	0.59
Average	3.1	3.2	3.4	3.5	3.2	3.3	3.4	3.2	3.0	3.0	3.2	0.17



* B1=Direction for reorganizing the national structure

** B2=Environmental issues

Figure 6.4. The Hierarchical Structure for the Capital Relocation Problem

6.7.4 Weights of the 18 Criteria by the 18 Council Members (Evaluators)

We employed a multi-stage use of an AHP¹⁵-like method for 18 of the 19 Council members (excluding the Chairperson). This went as follows. At the first

stage, the council member assigned their weights on criteria either by subjective judgments (not by AHP) or by AHP. In the latter case, we allowed incomplete paired comparisons so that the council member could skip the comparison when he or she possessed little or no confidence in comparing the criteria. At the end of the first stage, we gathered 18 sets of weights on 18 criteria. We then went on to the second stage, where we first showed the distribution of weight scores to the Council. Each member thus knew where he or she was located in the distribution. Taking account of this distribution, each member had the chance to alter his/her decision on criteria weights. Note that this is a form of the Delphi method.¹⁶ We continued this process until convergence was obtained. Actually, this process converged at the third stage. The resulting statistics for this evaluations are exhibited in Table 6.10. Each member had 100 points allotted for the evaluation to be divided and assigned to the 18 criteria according to his/her individual judgments. Observe that criteria C5 (Speedy response to

Table 6.10. Statistics of Weights assigned the 18 Criteria (C1-C18) by 18 Council Members

	Average	Median	Max	Min
C1	5.64822	5.45588	10.00000	1.24549
C2	4.72971	4.72222	10.00000	1.00000
C3	5.14061	5.27778	8.00000	1.55039
C4	6.79305	5.97666	16.42193	2.00000
C5	9.33093	8.49438	16.47018	5.00000
C6	5.6461	5.68774	9.00000	1.70505
C7	6.11132	5.68774	13.0372	2.81531
C8	6.13195	5.87875	10.00000	2.52025
C9	4.41615	3.89019	9.66938	1.16727
C10	8.46151	7.54204	19.92784	3.00000
C11	5.85123	5.44406	10.31643	2.11388
C12	6.97198	7.44607	15.50388	1.37530
C13	4.57878	4.34375	7.75194	1.94630
C14	4.01225	4.59094	5.75751	1.55039
C15	4.45201	4.57263	8.00000	1.00000
C16	4.27136	4.16636	7.17284	1.37530
C17	4.17737	3.62176	7.97114	2.00000
C18	3.27548	2.89578	6.42064	1.09375

a large-scale disaster), C10 (Safety in the event of an earthquake disaster), C12 (Feasibility of the smooth acquisition of land) and C4 (Ease of correspondence with a new information network) have high scores, on average. However, a large variation in evaluations exists, as can be seen from the “Max” and “Min” columns. This reflects the existence of significant disagreement among the Council members. We note that the importance of the criterion B1 “Direction

for reorganizing the national structure" is 10.378532 (the sum of C1 and C2) on average and that of B2 "Environmental issues" is 7.45285 (the sum of C17 and C18). Also the three top-criteria have the following weights: (Influence on the future of the country) = 31.64252, (Conditions for establishing the capital city functions) = 36.61826 and (Conditions for suitability of the new location) = 31.73923. Eventually, the top three criteria were faced to have approximately equal weights.

6.7.5 Decision Analyses using Averages and Medians

By using the averages and medians of the weights in Table 6.10, we obtained the corresponding scores for each site, as exhibited in Table 6.11. This table suggests Sites C (Tochigi), D (Tochigi/Fukushima) and G (Gifu/Aichi) as promising candidates.

Table 6.11. Averages and Medians of Scores of the 10 Sites

	A	B	C	D	E	F	G	H	I	J
Average	321	325	344	351	333	316	340	310	299	298
Median	308	310	328	334	317	302	328	297	287	287

6.7.6 Decision Analyses using the Assurance Region Model

However, using the average suggests that only one "virtual" evaluator was "representative" of all members judgments. Thus, the variety of opinions across evaluators is not taken into account. Given the degree of scatter exhibited in Table 6.10, the use of such an "average" or "median" of weights must be employed cautiously from a consensus-making point of view.

Another way to look at the above approach is that the weights are common to all sites. We may call this a "fixed weight" approach, as contrasted with the following "variable weight" structure.

Given the score matrix $S = (S_{ij})$, we evaluate the total score of site $j = j_o$ using a weighted sum of S_{ij_o} as

$$\theta_{j_o} = \sum_{i=1}^{18} u_i S_{ij_o}, \quad (6.62)$$

with a nonnegative weight set (u_i) . We assume that the weights can vary from site to site in accordance with the principle we choose for characterizing the sites.

Furthermore, the weights should reflect all evaluators preferences regarding the criteria. This can be represented by a version of the assurance region (AR)

model proposed by Thompson *et al.* (1986). For every pair (i, j) of criteria, the ratio u_i/u_j must be bounded by L_{ij} and U_{ij} as follows,

$$L_{ij} \leq u_i/u_j \leq U_{ij}, \quad (6.63)$$

where the bounds are calculated by using evaluator k 's weights (W_{ki}) on criterion i as

$$L_{ij} = \min_{k=1, \dots, 18} \frac{W_{ki}}{W_{kj}}, \quad U_{ij} = \max_{k=1, \dots, 18} \frac{W_{ki}}{W_{kj}}. \quad (6.64)$$

We now turn to the evaluation of candidate sites by means of the assurance region (AR) model of DEA. For this purpose, we employ two extreme cases as presented below.

6.7.7 Evaluation of “Positive” of Each Site

In order to evaluate the positives of site j_o , we choose the weights (u_i) in (6.62) so that they maximize θ_{j_o} under the condition that the same weights are applied in evaluating all other sites, and that the objective site should be compared relative to them. This principle is in accordance with that of DEA and can be formulated as follows:

$$\max \quad \theta_{j_o} = \sum_{i=1}^{18} u_i S_{ij_o} \quad (6.65)$$

$$\begin{aligned} \text{subject to} \quad & \sum_i u_i S_{ij} \leq 1 \quad (\forall j) \\ & L_{ij} \leq u_i/u_j \leq U_{ij} \quad (\forall (i, j)) \\ & u_i \geq 0, \quad (\forall i) \end{aligned} \quad (6.66)$$

where L_{ij} and U_{ij} are given by (6.64). The optimal score $\theta_{j_o}^*$ indicates the relative distance from the efficient frontier. The lower the score, the weaker the “positive” of the site.

It should be noted that DEA is here directed toward “effectiveness” rather than “efficiency” since we do not deal with resource utilization, as required for evaluating efficiency. See Section 3.9 in Chapter 3. Our concern is with achieving the already stated (or prescribed) goals. The initial goals, stated broadly, are made sufficiently precise with accompanying criteria for evaluation so that (a) proposed actions can be evaluated more accurately, and (b) once the proposals are implemented, any accomplishments (or lack thereof) can be subsequently identified and evaluated.

6.7.8 Evaluation of “Negative” of Each Site

In the above evaluations, each site was compared with the best performers. We will call this evaluation scheme as “positives,” in that we observe the candidate

from the positive side. Turning to the opposite side, we would like to evaluate the candidate sites from the worst side. For this purpose, we try seek the “worst” weights in the sense that the objective function in (6.65) is minimized. Thus, this principle can be formulated as follow:

$$\min \quad \phi_{j_o} = \sum_i u_i S_{ij_o} \quad (6.67)$$

$$\text{subject to} \quad \sum_i u_i S_{ij} \geq 1 \quad (\forall j) \quad (6.68)$$

$$L_{ij} \leq u_i / u_j \leq U_{ij} \quad (\forall (i, j)) \quad (6.69)$$

$$u_i \geq 0. \quad (\forall i) \quad (6.70)$$

By dint of the reversed inequality in (6.68), the optimal ϕ_{j_o} satisfies $\phi_{j_o}^* \geq 1$. If $\phi_{j_o}^* = 1$, then the site belongs to the worst performers group; otherwise, if $\phi_{j_o}^* > 1$, it rates higher than the worst performers group. Each site is compared with these worst performers and is gauged by its efficiency “negatives” as the ratio of distances from the “worst” frontiers in the same way as in ordinary DEA. (Yamada *et al.*¹⁷ named this worst side approach “Inverted DEA.”)

In order to make straightforward comparisons of the “negatives” and “positives” case scores, we invert $\phi_{j_o}^*$ as follows

$$\tau_{j_o}^* = 1/\phi_{j_o}^* \quad (6.71)$$

and call it the “negatives” score.

6.7.9 Uses of “Positive” and “Negative” Scores

Since the number of candidate sites (10) is smaller than the number of criteria (18), we have a shortfall in the number of degrees of freedom for discriminating efficiency among the 10 candidates even if we employ the AR models.

First, we estimated the lower/upper bounds, L_{ij} and U_{ij} , respectively, on the ratio of criteria i and j by (6.64). If sufficient discrimination among the candidates both in “positives” and “negatives” cannot be observed, we apply the following deletion processes. We delete the k s giving the *min* and *max* ratios in (6.64), and estimate the *min* and *max* again using (6.64). Thus, we remove two extreme ratios at each trial. Our experiments showed that after six deletions, using the remaining six (=18-12) ratios, we reached a sufficient and stationary evaluation of sites, as depicted in Figure 6.5. From this figure, we observe that two Sites, D (positive=1, negative=0.894) and G (positive=1, negative=0.898) are excellent in both “positives” and “negatives.” The Site C is somewhat behind D and G.

6.7.10 Decision by the Council

From the above analyses, the Council acknowledged Sites C, D and G as the most promising candidates for their selection problem. However, since Site C (Tochigi district) is a part of Site D (Tochigi/Fukushima district), the Council decided to recommend Sites D and G as final. They also recommended

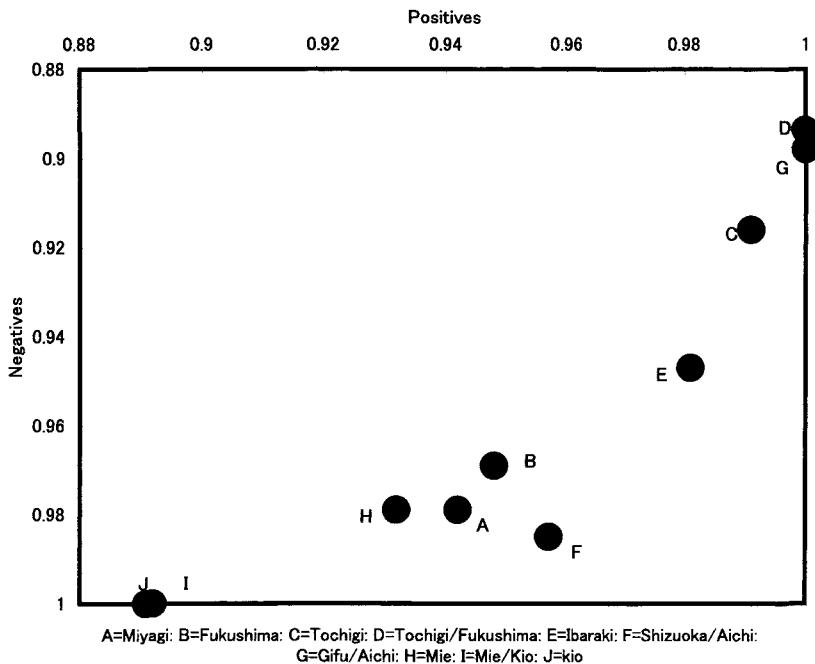


Figure 6.5. Positives and Negatives of the 10 Sites

Site I (Mie/Kio district) as the runner-up provided that this district would be powered by a high-speed transportation network. The report was delivered to Prime Minister who transmitted it to the Diet where it still being considered with each of the three sites enthusiastically advertising their superiority.

6.7.11 Concluding Remarks

We have presented a report on the site selection process for relocating Government Agencies out of Tokyo, putting emphasis on the methodological aspects.

It is hoped that our methodologies will be utilized in a successful execution of this centennial project with satisfactory results.

We summarize key characteristics of the proposed methods below.

These methods assume the following two conditions:

1. Each site has been evaluated numerically with respect to the set of chosen criteria. These evaluations may be made objectively by using measurable yardsticks, or they may be done subjectively using expert knowledge.
2. Each evaluator can determine his/her own judgment on the relative importance of the criteria. For this purpose, AHP or direct subjective judgments may be utilized.

In cases where these conditions are satisfied, the proposed methods rank the candidate sites to identify a consensus within the evaluator group. Results obtained by the assurance region model have, in particular, several merits for both candidates and evaluators, as follows:

- **For Candidate Sites:** The results are acceptable to the candidate sites in the sense that the most preferable weights for the site are assigned within the allowable bounds of the evaluators. The optimal weights vary from site to site in that the best set of relative weights is assigned to the site. In a similar way, the relative weaknesses of each site can also be evaluated. These two measures are then utilized to characterize the candidate sites.
- **For Evaluators:** Each evaluator can be assured that his/her judgments on the criteria are taken into account and that the ratios of every pair of weights fall within his/her allowable range. Although in this case several evaluators ratios were excluded for discrimination purposes, this approach is more reasonable and acceptable than using the average (or median) weights of all evaluators, especially when there is a relatively high degree of scatter to consider.

The proposed approaches can also be applied to many other societal problems that involve group decision making.

6.8 SUMMARY OF CHAPTER 6

In this chapter, we introduced the assurance region and cone-ratio methods for combining subjective and expert evaluations with the more objective methods of DEA.

1. Usually expressed in the form of lower and upper bounds, the assurance region method puts constraints on the ratio of input (output) weights or multiplier values. This helps to get rid of zero weights which frequently appear in solution to DEA models. The thus evaluated efficiency score generally drops from its initial (unconstrained) value. Careful choice of the lower and upper bounds is recommended.
2. Not covered in this chapter is the topic of “linked constraints” in which conditions on input and output multipliers are linked. See Problem 6.3.
3. The cone-ratio method confines the feasible region of virtual multipliers v , u , to a convex cone generated by admissible directions. Formulated as a “cone ratio envelopment” this method can be regarded as a generalization of the assurance region approach.
4. Example applications were used to illustrate uses of both of the “assurance region” and “cone ratio envelopment” approaches.

6.9 NOTES AND SELECTED BIBLIOGRAPHY

The assurance region (AR) method was developed by Thompson, Singleton, Thrall and Smith (1986).¹⁸ They used DEA to analyze six Texas sites for location of a high energy physics lab (called the SSC = Supercolliding Super Conductor) which was directed to advancing fundamental knowledge in physics. Five of the six sites were DEA efficient. This was not satisfactory so they then used survey data and expert opinion to specify bounds for the virtual multipliers. The AR method identified only one efficient DMU for the location of SSC and this candidate site was selected by Texas and won in a national competition conducted by the U.S. Department of Energy in 1988 as the location for the SSC. Further development of the AR method can be found in Thompson, Langemeir, Lee, Lee and Thrall (1990)¹⁹ and Roll and Golany (1993)²⁰. See also Dyson and Thanassoulis (1988).²¹ The use of DEA for site selection has been fully developed in Athanassopoulos and Storbeck (1995)²² and Desai, Haynes and Storbeck (1994).²³

The cone-ratio (CR) method was developed by Sun (1988)²⁴, Charnes, Cooper, Wei and Huang (1989)²⁵ and Charnes, Cooper, Huang and Sun (1990),²⁶ which they applied to large U.S. commercial banks. See also Brockett, Charnes, Cooper, Huang and Sun (1997)²⁷ which includes an Appendix on how to derive original (untransformed) data and solutions.

6.10 RELATED DEA-SOLVER MODELS FOR CHAPTER 6

AR-I-C (Input-oriented Assurance Region model under the Constant returns-to-scale assumption).

This code solves the AR problem expressed by (6.11)-(6.14). The main body of the data set, (X, Y) , for this model is put on an Excel worksheet in the same way as the CCR model. The assurance region (AR) is augmented by adding AR constraints such as (6.1) under the main data set separated by one blank row. See Figure B.2.1 in Section B.5.2 for a sample format and refer to the explanations above the figure. The main results will be obtained in the following worksheets.

- The efficiency score θ^* and the reference set (λ^*) are recorded in “Score.”
- The optimal (dual) multipliers v^*, u^* in (6.6) and the weighted data $\{x_{ij}v_i^*\}$, $\{y_{rj}u_r^*\}$ are displayed in “Weight” and “WeightedData.” They satisfy the assurance region constraints.
- The projection onto efficient frontiers by (6.17) and (6.18) is stored in “Projection.”
- Slacks s^{-*} and s^{+*} defined by (6.15) and (6.16) are included in “Slack.”

AR-I-V (Input-oriented assurance region model under the variable returns-to-scale assumption).

This code solves the AR model under the added constraint $\sum_{j=1}^n \lambda_j = 1$. In addition to results similar to those in AR-I-C, this code identifies the

returns-to-scale characteristics of each DMU under the AR constraints. This identification is based on Tone (2001).²⁸

- The returns to scale of each DMU is recorded in the worksheet “RTS.” This RTS is characterized under the AR environments. Thus it may differ from the one obtained by the BCC-I model, which has no constraints on the multipliers v , u , and hence the structure of the supporting hyperplanes changes from that of the BCC model. The RTS of inefficient DMUs is identified as that of the projected DMU on the efficient frontier in the input-oriented manner.

AR-I-GRS (Input-oriented assurance region model under the general returns-to-scale assumption).

This code solves the AR model under the added constraint $L \leq \sum_{j=1}^n \lambda_j \leq U$. The lower bound $L(\leq 1)$ and upper bound $U(\geq 1)$ must be supplied through keyboard. The defaults are $L = 0.8$ and $U = 1.2$.

AR-O-C (Output-oriented assurance region model under the constant returns-to-scale assumption).

This code is the output-oriented version of AR-I-C. We exhibit the optimal efficiency score θ^* as the inverse of the optimal output expansion rate so that it is less than or equal to one. This will make the comparisons with the input-oriented case straightforward.

AR-O-V(GRS) (Output-oriented assurance region model under the variable (general) returns-to-scale assumption).

The code AR-O-V also identifies the RTS of each DMU in the output-oriented manner.

ARG-I(O)-C(V or GRS) (Input (Output)-oriented Assurance Region Global models under the constant (variable or general) returns-to-scale assumption.)

This code solves the ARG problem expressed by (6.29)-(6.32). The main body of the data set, (X, Y) , for this model is put on an Excel worksheet in the same way as the CCR model. The assurance region global (ARG) is augmented by adding ARG constraints under the main data set separated by one blank row. See Figure B.2.1 in Section B.5.2 for a sample format and refer to the explanations above the figure. The main results will be obtained in the same scheme as the AR models.

6.11 PROBLEM SUPPLEMENT FOR CHAPTER 6

Problem 6.1

In addition to “cone-ratio envelopment” can you identify other approaches to data transformations that are used to effect evaluations?

Suggested Response : One such possibility is the use of data from benchmark

firms to obtain ratios that are then used to evaluate performances of other firms relative to these benchmarks. Other examples are (1) the use of index numbers of prices as divisors to obtain measures of “real” vs. “nominal” (dollar) performances of productivity and (2) the use of discount rates to replace future income streams with a single “present value.”

Problem 6.2

Part(1) : Show that (6.59) can lead to trouble in the form of infinite solutions to (6.57) when the slacks are not constrained to be nonnegative.

Part(2) : Analyze the dual to this problem and use the analysis to show how such troubles can be avoided.

Part(3) : Show that such troubles do not occur when the transformed data are used as in (6.58).

Suggested Response :

Part(1): Consider the following CCR formulation in terms of the original data,

$$\begin{aligned} \min \quad & \theta \\ \text{subject to} \quad & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + \tilde{s}_i^-, \quad i = 1, \dots, m \\ & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - \tilde{s}_i^+, \quad r = 1, \dots, s \end{aligned}$$

where the $\lambda_j \geq 0 \forall j$ but the slacks not otherwise constrained. Setting all $\lambda_j = 0$ we can choose $-\tilde{s}_r^+ = y_{ro} \forall r$ and then let $\tilde{s}_i^- = \theta x_{io}$ go to negative infinity in both \tilde{s}_i^- and θ .

Part(2): Consider the dual to this last problem in the form

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} \\ \text{subject to} \quad & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n \\ & \sum_{i=1}^m v_i x_{io} = 1 \\ & u_r, v_i = 0 \quad \forall i, r. \end{aligned}$$

This has no solution because the condition $\sum_{i=1}^m v_i x_{io} = 1$ is not compatible with $v_i = 0, \forall i$. However, changing this to $\tilde{s}_i^- \geq 0$ for some of the primal (= envelopment model) slacks removes this difficulty. Both problems will then have solutions so that, by the dual theorem of linear programming, they will have finite and equal optima. \square

(Note, however, that it is only the $v_i^*, u_r^* > 0$ that can serve as multipliers. Hence only the slacks associated with these variables can directly contribute to the optimum objective. The other variables can contribute only indirectly.)

Part(3): Write the dual to (6.57) — and (6.58) — as

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r \bar{y}_{ro} \\ \text{subject to} \quad & \sum_{r=1}^s u_r \bar{y}_{rj} - \sum_{i=1}^m v_i \bar{x}_{ij} \leq 0, \quad j = 1, \dots, n \\ & \sum_{i=1}^m v_i \bar{x}_{io} = 1 \\ & u_r, v_i \geq 0 \quad \forall i, r. \end{aligned}$$

Choosing $v_k = 1/\bar{x}_{ko}$ and all other variables zero satisfies all constraints. Hence we have exhibited a solution. Moreover, the constraints require

$$\max \sum_{r=1}^s u_r \bar{y}_{ro} \leq \sum_{i=1}^m v_i \bar{x}_{io} = 1$$

so the solutions are bounded. The dual theorem of linear programming asserts that if one member of a dual pair has a finite optimum then the other member will have a finite and equal optimum. Thus infeasible solution possibilities in the envelopment model and the non-existence of solutions in the dual (multiplier) model are both avoided. \square

Problem 6.3

Background : The assurance region constraints in this chapter were restricted to separate conditions on the inputs and the outputs, respectively. The matrix D in (6.55) is then interpretable in terms of two convex cones, one in the input and the other in the output space. Extending this matrix to

$$\begin{pmatrix} A & O \\ O & B \\ F_1 & F_2 \end{pmatrix}$$

eliminates these characterizations by providing a linkage between the constraints associated with F_1 and F_2 . A use of such “linkages” in the form of assurance region constraints may be needed to deal with conditions on profits, or other relations between inputs and outputs — as noted in R.G. Thompson, P.S. Dharmpala and R.M. Thrall, “Linked-cone DEA Profit Ratios and Technical Inefficiencies with Applications to Illinois Coal Mines,” *International Journal of Production Economics*, 39, 1995, pp.99-115.

As these authors note in dealing with such “linkage constraints,” especially in an assurance region approach to profit conditions, constraints like the following

can have untoward effects in limiting profit evaluations,

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n.$$

Assignment : Discuss the effects of such constraints on profits and suggest ways to handle the resulting problems.

Suggested Response : Such constraints are limitational because one wants results like

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} > 0,$$

where the u_r and v_i are to be interpreted as prices and unit costs. Thompson, Dharmapala and Thrall suggest eliminating all such constraints, (and the data they contain,) in favor of formulations like the following.

$$\begin{aligned} \max \quad & \mathbf{u} \mathbf{y}_o \\ \text{subject to} \quad & \mathbf{v} \mathbf{x}_o = 1 \\ & \begin{pmatrix} A & O \\ O & B \\ F_1 & F_2 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \leq \mathbf{0}, \end{aligned}$$

in which the condition $\mathbf{v} \mathbf{x}_o = 1$ is retained but the constraints employing the other data are discarded.

The idea is to use the above formulation as a means of determining whether profits are possible in the following manner.

Step 1 Determine $\max \mathbf{u} \mathbf{y}_o = MPR$.

Step 2 Determine $\min \mathbf{u} \mathbf{y}_o = mPR$.

Then use the following conditions.

- $MPR > 1$ indicates a positive profit potential.
- $MPR < 1$ only losses are possible.
- $mPR > 1$ positive profits are assured.
- $mPR < 1$ indicates a loss potential.

The following single input - single output example is provided by Thompson, Dharmapala and Thrall. Let $y_o = 1$ and $x_o = 1$ and require $3v \leq u, u \leq 5v$. The above formulation then becomes

$$\begin{aligned} \max \quad & u \\ \text{subject to} \quad & v = 1 \\ & 3v - u \leq 0 \\ & -5v + u \leq 0 \\ & u, v \geq 0. \end{aligned}$$

Step 1 gives $MPR = 5 > 1$. Reorienting the objective, as required in Step 2, gives $mPR = 3 > 1$ so this DMU_o is assured of a profit.

Comment : Numerous DEA studies have shown that profitable firms, and even the most profitable firms, are not necessarily Pareto-Koopmans efficient, i.e., in the sense of Definition 3.3 it is possible to improve some inputs and outputs without worsening other inputs and outputs. Hence this approach to the use of “linked” assurance region constraints can help to identify where attention can be advantageously directed.

Problem 6.4

Prove that if the linear program (DAR_o) associated with the assurance region method in Section 6.2 is feasible with a positive optimum ($\theta^* > 0$), then the reference set defined by $E_o = \{j | \lambda_j^* > 0\}$ ($j \in \{1, \dots, n\}$) is not empty.

Suggested Answer : An optimal solution $(\theta^*, \lambda^*, \pi^*, \tau^*, s^{-*}, s^{+*})$ of (DAR_o) satisfies:

$$\theta^* x_o = X \lambda^* - P \pi^* + s^{-*}.$$

Let an optimal solution of (AR_o) be (v^*, u^*) . By multiplying the above equation by v^* , we have:

$$\theta^* v^* x_o = v^* X \lambda^* - v^* P \pi^* + v^* s^{-*}.$$

From a constraint of (AR_o), it holds $v^* x_o = 1$ and the complementarity relations assert $v^* P \pi^* = 0$ and $v^* s^{-*} = 0$. Hence we have:

$$\theta^* = v^* X \lambda^* > 0.$$

Therefore λ^* must be semipositive, i.e., $\lambda^* \geq 0$ and $\lambda^* \neq 0$. Hence E_o is not empty. \square

Notes

1. R.G. Thompson, F.D. Singleton, Jr., R.M. Thrall and B.A. Smith (1986), “Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas,” *Interfaces* 16, pp. 35-49. See also R.G. Dyson and E.Thanassoulis (1988), “Reducing Weight Flexibility in Data Envelopment Analysis,” *Journal of the Operational Research Society*, 39, pp.563-576. Finally, see Notes and Selected Bibliography in Section 6.9 of this chapter for references on uses of DEA for site selection.

2. A. Charnes, W.W. Cooper, Z.M. Huang and D.B. Sun (1990), “Polyhedral Cone-Ratio DEA Models with an Illustrative Application to Large Commercial Banks,” *Journal of Econometrics* 46, pp.73-91. For a treatment that studies this approach as an alternative to the more rigid approach to risk evaluation under the “Basel Agreement” for controlling risks in bank portfolios see P.L. Brockett, A. Charnes, W.W. Cooper, Z.M. Huang and D.B. Sun, “Data Transformations in DEA Cone-Ratio Approaches for Monitoring Bank Performance,” *European Journal of Operational Research*, 98, 1997, pp.250-268.

3. Y. Roll, W.D. Cook and B. Golany (1991), “Controlling Factor Weights in Data Envelopment Analysis,” *IIE Transactions* 23, pp.2-9.

4. This usage is due to Vilfredo Pareto. See the references in Chapter 3, Section 3.10: "Notes and Selected Bibliography."

5. See Definition 3.1.

6. This problem can be solved by using "AR-I-C" (the input-oriented assurance region model under the constant returns-to-scale assumption) in DEA-Solver.

7. R. Allen, A. Athanassopoulos, R.G. Dyson and E. Thanassoulis (1997), "Weights restrictions and value judgements in data envelopment analysis," *Annals of Operations Research* 73, pp.13-34.

8. P.L. Brockett, A. Charnes, W.W. Cooper, Z. Huang and D.B. Sun (1997), "Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performance," *European Journal of Operational Research* 98, pp.250-268.

9. For detailed discussions see the references cited in Brockett *et al.*(1997).

10. In response to this trend, Barr, Seiford, and Siems with the Federal Reserve Bank of Dallas developed a bank failure prediction model based on DEA which outperforms all other failure prediction models in the banking literature. See R.S. Barr, L.M. Seiford and T.F. Siems (1994), "Forcasting Bank Failure: A Non-Parametric Frontier Estimation Approach," *Recherches Economiques de Louvain* 60, pp.417-429 and R.S. Barr, L.M. Seiford and T.F. Siems (1993), "An Envelopment-Analysis Approach to Measuring the Managerial Efficiency of Banks," *Annals of Operations Research*, 45, pp.1-19 for details.

11. See the definition and discussion of "Reserve (=Allowance) for Bad Debts" on page 433 in *Kohler's Dictionary for Accountants*, 6th Edition (Englewood Cliffs, N.J., Prentice-Hall Inc., 1983.)

12. R.M. Nun (1989), "Bank Failure: The Management Factor" (Austin, TX., Texas Department of Banking).

13. See A. Ben Israel and T.N. Greville, *Generalized Inverses* (New York, John Wiley & Sons Inc., 1974).

14. Y. Takamura and K. Tone (2003), "A Comparative Site Evaluation Study for Relocating Japanese Government Agencies out of Tokyo," *Socio-Economic Planning Sciences* 37, pp.85-102.

15. AHP (Analytic Hierarchy Process) was invented by T.L. Saaty (1980), *Analytic Hierarchy Process*, New York: McGraw-Hill. See also K. Tone (1989), "A Comparative Study on AHP and DEA," *International Journal of Policy and Information* 13, pp.57-63.

16. See J.A. Dewar and J.A. Friel (2001), "Delphi Method," in S.I. Gass and C.M. Harris, eds., *Encyclopedia of Operations Research and Management Science* (Norwell, Mass., Kluwer Academic Publishers) pp.208-209.

17. Y. Yamada, T. Matsui and M. Sugiyama (1994), "An Inefficiency Measurement Method for Management Systems," *Journal of the Operations Research Society of Japan* 37, 2, pp.158-167.

18. See the Note 1 reference.

19. R.G. Thompson, L.N Langemeir, C. Lee, E. Lee and R.M. Thrall (1990), "The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming," *Journal of Econometrics* 46, pp.93-108.

20. Y. Roll and B. Golany (1993), "Alternate Methods of Treating Factor Weights in DEA," *OMEGA* 21, pp.99-109.

21. R.G. Dyson and E. Thanassoulis (1988), "Reducing Weight Flexibility in Data Envelopment Analysis," *Journal of the Operational Research Society* 39, pp.563-576.

22. A.D. Athanassopoulos and J.E. Storbeck (1995), "Non-Parametric Models for Spatial Efficiency," *The Journal of Productivity Analysis* 6, pp.225-245.

23. A. Desai, K. Haynes and J.E. Storbeck "A Spatial Efficiency Framework for the Support of Locational Decisions," in *Data Envelopment Analysis: Theory, Methodology, and Applications*, A. Charnes, W. W. Cooper, Arie Y. Lewin, and Lawrence M. Seiford (editors), Kluwer Academic Publishers, Boston, 1994.

24. D.B. Sun (1988), "Evaluation of Managerial Performance in Large Commercial Banks by Data Envelopment Analysis," Ph.D. dissertation, Graduate School of Business, University of Texas, Austin, TX.
25. A. Charnes, W.W. Cooper, Q.L. Wei and Z.M. Huang (1989), "Cone Ratio Data Envelopment Analysis and Multi-objective Programming," *International Journal of Systems Science* 20, pp.1099-1118.
26. See the Note 2 reference.
27. See the Note 8 reference.
28. K. Tone (2001), "On Returns to Scale under Weight Restrictions in Data Envelopment Analysis," *Journal of Productivity Analysis* 16, pp.31-47.

7

NON-DISCRETIONARY AND CATEGORICAL VARIABLES

7.1 INTRODUCTION

Section 3.10 in Chapter 3 introduced the topic of non-discretionary variables and applied it to examples drawn from a study of Texas schools. In that study it was necessary to allow for variations in “minority,” “economically disadvantaged” and “low English proficiency” students who had to be dealt with in different schools. These input variables were “non-discretionary.” That is, they could not be varied at the discretion of individual school managers but nevertheless needed to be taken into account in arriving at relative efficiency evaluations.

The approach taken in Chapter 3, as adapted from Banker and Morey ¹ (1986), took form in the following model:

$$\begin{aligned} \min \quad & \theta - \varepsilon \left(\sum_{i \in D} s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{subject to} \quad & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in D \\ & x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in ND \end{aligned} \tag{7.1}$$

$$y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s.$$

where all variables (except θ) are constrained to be nonnegative. See (3.85)-(3.88).

Here the symbol D and ND refer to “Discretionary” and “Non-Discretionary,” respectively. Focussed on the inputs, as in (7.1) above, such non-discretionary variables may also be extended to outputs.

This topic can also be treated in different ways. For instance, Charnes *et al.* (1987)² extended the Additive model in order to accommodate non-discretionary variables in the following form.

$$\begin{aligned} \max \quad & \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+ \\ \text{subject to} \quad & \sum_{j=1}^n x_{ij} \lambda_j + s_i^- = x_{io}, \quad i = 1, \dots, m \\ & \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ = y_{ro}, \quad r = 1, \dots, s \\ & s_i^- \leq \beta_i x_{io}, \quad i = 1, \dots, m \\ & s_r^+ \leq \gamma_r y_{ro}, \quad r = 1, \dots, s \end{aligned} \tag{7.2}$$

where the β_i, γ_r represent parameters (to be prescribed) and all variables are constrained to be nonnegative.

Assigning values from 0 to 1 accords different degrees of discretion to input i with $\beta_i = 0$ characterizing this input as completely non-discretionary and $\beta_i = 1$ changing the characterization to completely discretionary. Similarly setting $\gamma_r = 0$ consigns output r to a fixed (non-discretionary) value while allowing $\gamma_r \rightarrow \infty$, or, equivalently, removing this constraint on s_r^+ allows its value to vary in a freely discretionary manner.

As can be seen, the latter model can be used to eliminate the nonzero slack that can appear with the Banker-Morey approach associated with (7.1). Such nonzero slack is omitted from the *measure* of efficiency in the B-M approach, of course, but it can be informative to, say, higher levels of management as impounded in the constraints. For instance, the ability to handle additional “economically disadvantaged” students noted in Table 3.7 could be used in redistricting this school to improve the performances of *other* schools in which this constraint is “tight” — as exhibited by zero slack and a positive value for the dual (multiplier) variable associated with this constraint in other schools.

This is not the end of the line, of course, so additional approaches to the treatment of non-discretionary variables will be undertaken in this chapter. There are related topics that will also be treated. One such topic involves the use of categorical (classificatory) variables which assign DMUs to different classes such as “large,” “small,” “medium,” etc.

Another such topic involves the introduction of constraints on the range of values allowable for the variables. The “assurance regions” and “cone-ratio envelopments” developed in Chapter 6 approached this topic in terms of the multiplier variables. Here the constraints will be imposed on variables in the envelopment models.

7.2 EXAMPLES

Examples from Figure 7.1 can help us understand the treatment in these two models. This figure geometrically portrays five DMUs which use two inputs

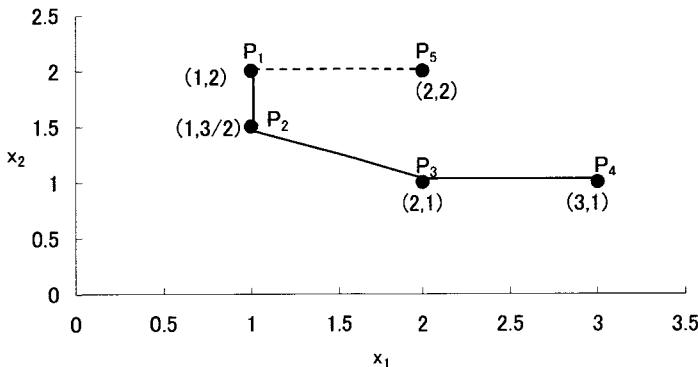


Figure 7.1. One Input Exogenously Fixed, or Non-Discretionary

to produce one unit of the same output. The points P_1, \dots, P_5 represent the similarly numbered DMUs. The first parenthesized number is the value for the input amount x_1 and the second number is the value of the input amount x_2 . We assume that the variable x_2 is non-discretionary. To evaluate P_5 via (7.1) we use

$$\begin{aligned} \min \quad & \theta - \varepsilon(s_1^- + s^+) \\ \text{subject to} \quad & 2\theta = \lambda_1 + \lambda_2 + 2\lambda_3 + 3\lambda_4 + 2\lambda_5 + s_1^- \\ & 2 = 2\lambda_1 + 3/2\lambda_2 + \lambda_3 + \lambda_4 + 2\lambda_5 + s_2^- \\ & 1 = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 - s^+ \end{aligned}$$

with all variables constrained to be nonnegative. Noting that s_2^- is not in the objective (where the efficiency evaluation measures appear) we find that the following solutions are alternative optima,

$$\begin{aligned} \theta^* &= 1/2, \quad \lambda_1^* = 1 \\ \theta^* &= 1/2, \quad \lambda_2^* = 1, \quad s_2^{-*} = 1/2 \end{aligned}$$

where, in each case, all other variables are zero.

Recourse to (7.2) replaces the preceding example with

$$\begin{aligned} \max \quad & s_1^- + s_2^- + s^+ \\ \text{subject to} \quad & \lambda_1 + \lambda_2 + 2\lambda_3 + 3\lambda_4 + 2\lambda_5 + s_1^- = 2 \\ & 2\lambda_1 + 3/2\lambda_2 + \lambda_3 + \lambda_4 + 2\lambda_5 + s_2^- = 2 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 - s^+ = 1 \\ & s_1^- \leq 2\beta_1 \\ & s_2^- \leq 2\beta_2 \\ & s^+ \leq 1\gamma \end{aligned}$$

where, again, all variables are also nonnegative. The second input is the only non-discretionary variable so we set

$$\beta_2 = 0, \beta_1 = 1, \text{ and } \gamma \rightarrow \infty.$$

This choice yields

$$\lambda_1^* = 1, s_1^{-*} = 1$$

and all other variables zero. Because s_2^- is restricted to a zero value in the constraint as well as in the objective, the preceding alternate optimum is eliminated and the solution is uniquely optimal.

When all variables are discretionary a DMU will be efficient if its associated vector of observations enters into an optimal basis with a positive coefficient. Such vectors are referred to as "active" members of the basis in order to distinguish them from other vectors which enter with zero coefficients because they are needed only to complete a basis and, hence, need not be efficient. See Problems 7.5 and 7.6, below.

As can be seen, P_1 has been elevated to the status of being efficient because it is an active member of the optimal basis used to evaluate P_5 . The value of $s_2^{-*} = 1/2$ in the alternate optimum signals that this efficient status might change when the variable $x_2 = 2$ is changed from discretionary to non-discretionary. This information can be of interest to, say, higher levels of management, as was noted in our discussion of Table 3.5 in Chapter 3.

A common thread in the approaches used in (7.1) and (7.2) is the fact that nonzero slacks which appear only in the constraints do not enter into the efficiency evaluations. Hence for CCR (or BCC) efficiency, as represented in (7.1), Definition 3.2 is sharpened to the following

Definition 7.1 *Full (CCR or BCC) efficiency is achieved for DMU_o if and only if both of the following conditions are satisfied*

- (i) $\theta^* = 1$
- (ii) All slacks in the objective are zero.

For the extended additive model, as represented in (7.2), Definition 4.3 is modified to

Definition 7.2 *All slacks at zero in the objective is a necessary and sufficient condition for full efficiency with (7.2).*

The refinements in these definitions were not needed when all variable values could be altered at discretion of the managers. The consequences of the presence of non-discretionary variables can be examined a bit further by evaluating P_1 with $x_2 = 2$ fixed at its observed value. Thus using (7.1) we obtain

$$\begin{aligned} \min \quad & \theta - \varepsilon(s_1^- + s^+) \\ \text{subject to} \quad & 1\theta = \lambda_1 + \lambda_2 + 2\lambda_3 + 3\lambda_4 + 2\lambda_5 + s_1^- \\ & 2 = 2\lambda_1 + 3/2\lambda_2 + 1\lambda_3 + 1\lambda_4 + 2\lambda_5 + s_2^- \\ & 1 = 1\lambda_1 + 1\lambda_2 + 1\lambda_3 + 1\lambda_4 + 1\lambda_5 - s^+ \end{aligned}$$

where all variables except θ are constrained to be nonnegative. This has alternate optima,

$$\begin{aligned} \theta^* &= 1, \quad \lambda_2^* = 1, \quad s_2^{-*} = 1/2 \\ \theta^* &= 1, \quad \lambda_1^* = 1 \end{aligned}$$

and all other variables zero. In either case P_1 is characterized as efficient because $\theta^* = 1$ and $s_2^{-*} = 1/2$ do not appear in the objective. See Definition 7.1.

7.3 NON-CONTROLLABLE, NON-DISCRETIONARY AND BOUNDED VARIABLE MODELS

Further discussion and development of the above examples are provided in the problems at the end of this chapter. Here we turn to an application which uses a different (but related) formulation and terminology as in the following formulations.

7.3.1 Non-controllable Variable (NCN) Model

$$(NCN) \quad \min_{\theta, \boldsymbol{\lambda}} \quad \theta \tag{7.3}$$

$$\text{subject to} \quad \theta \mathbf{x}_o^C \geq X^C \boldsymbol{\lambda}$$

$$\mathbf{y}_o^C \leq Y^C \boldsymbol{\lambda}$$

$$\mathbf{x}_o^N = X^N \boldsymbol{\lambda} \tag{7.4}$$

$$\mathbf{y}_o^N = Y^N \boldsymbol{\lambda} \tag{7.5}$$

$$L \leq \mathbf{e} \boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq 0.$$

Here we are using a matrix-vector formulation in which X^C, Y^C refer to matrices of “controllable” variables and $\mathbf{x}_o^C, \mathbf{y}_o^C$ refer to the corresponding vectors of observed values for the DMU_o being evaluated. The matrices X^N, Y^N , on the other hand, refer to data on the non-controllable variables that are to be evaluated relative to the vectors $\mathbf{x}_o^N, \mathbf{y}_o^N$ for this same DMU_o. Finally the last constraint imposes an upper bound, U , and a lower bound, L , on the variable choices with $\mathbf{e} \boldsymbol{\lambda} = \sum_{j=1}^n \lambda_j$ in conformance with (5.37) in Chapter 5.

To simplify matters in the additions and extensions we wish to make, we have not included slacks in the objective of (7.3). For the non-controllable variables this conforms to the definition we have now provided. We have, however, also omitted the slacks for the non-controllable variables so this needs to be allowed for in the application we now introduce.

7.3.2 An Example of a Non-Controllable Variable

Table 7.1 shows the data for public libraries in the 23 Wards of the Tokyo Metropolitan Area in 1986. As the measurement items of efficiency we use the floor area (unit=1000m²), the number of books (unit=1000), staffs and the populations of wards (unit=1000) as inputs and the number of registered residents (unit=1000) and borrowed books (unit=1000) as outputs. Using these data, we evaluated the relative efficiency using both the input-oriented CCR model and the CCR model with population as a non-controllable variable (NCN). Table 7.2 depicts the results.

Table 7.1. Data for Public Libraries in Tokyo

No.	Ward	INPUT			OUTPUT	
		Area (1000m ²)	Books (1000)	Staff	Populat. (1000)	Regist. (1000)
L1	Chiyoda	2.249	163.523	26	49.196	5.561
L2	Chūo	4.617	338.671	30	78.599	18.106
L3	Taito	3.873	281.655	51	176.381	16.498
L4	Arakawa	5.541	400.993	78	189.397	30.810
L5	Minato	11.381	363.116	69	192.235	57.279
L6	Bunkyo	10.086	541.658	114	194.091	66.137
L7	Sumida	5.434	508.141	61	228.535	35.295
L8	Shibuya	7.524	338.804	74	238.691	33.188
L9	Meguro	5.077	511.467	84	267.385	65.391
L10	Toshima	7.029	393.815	68	277.402	41.197
L11	Shinjuku	11.121	509.682	96	330.609	47.032
L12	Nakano	7.072	527.457	92	332.609	56.064
L13	Shinagawa	9.348	601.594	127	356.504	69.536
L14	Kita	7.781	528.799	96	365.844	37.467
L15	Kōto	6.235	394.158	77	389.894	57.727
L16	Katsushika	10.593	515.624	101	417.513	46.160
L17	Itabashi	10.866	566.708	118	503.914	102.967
L18	Edogawa	6.500	467.617	74	517.318	47.236
L19	Suginami	11.469	768.484	103	537.746	84.510
L20	Nerima	10.868	669.996	107	590.601	69.576
L21	Adachi	10.717	844.949	120	622.550	89.401
L22	Ōta	19.716	1258.981	242	660.164	97.941
L23	Setagaya	10.888	1148.863	202	808.369	191.166

In evaluating the efficiency of a library, the population of the area is an important (input) factor. If we apply the input-oriented CCR model for this evaluation, the CCR-projection by (3.22) scales down all inputs by multiplying the optimal score $\theta^*(\leq 1)$ and further deletes slacks if any. “Population” is not an exception and hence the “CCR Projection” listed in Table 7.2 resulted in reductions of the population for inefficient libraries. In other words, the CCR score θ^* is evaluated under the assumption that it is possible to scale down radially (proportionally) all inputs so long as the reduced inputs remain in the production possibility set. However, “Population” is non-controllable and so we apply the model expressed by (7.4) for this NCN variable in order to evaluate the efficiency of libraries. As exhibited under CCR Score and NCN Score in Table 7.2, the latter differs from the CCR score and this shows the influence of this constraint (hence a modification of the production possibility set). Finally, as intended, the NCN-projection of population is the same as the original data listed under Data Population.

Table 7.2. Efficiency of Libraries by CCR and NCN

No	CCR Score	NCN Score	Data Population	CCR Projection	NCN Projection
L1	0.35	0.301	49.196	17.219	49.196
L2	0.792	0.643	78.599	62.237	78.599
L3	0.573	0.651	176.381	101.119	176.380
L4	0.719	0.618	189.397	136.116	189.397
L5	1	1	192.235	192.235	192.235
L6	1	1	194.091	194.091	194.091
L7	0.697	0.71	228.535	159.230	228.535
L8	0.58	0.7	238.691	138.520	238.691
L9	1	1	267.385	267.385	267.385
L10	0.705	0.793	277.402	195.313	277.402
L11	0.569	0.65	330.609	188.088	330.609
L12	0.758	0.773	332.609	252.235	332.609
L13	0.747	0.687	356.504	266.485	356.504
L14	0.722	0.8	365.844	263.972	365.844
L15	0.844	1	389.894	262.813	389.894
L16	0.582	0.766	417.513	211.251	417.513
L17	1	1	503.914	503.914	503.914
L18	0.787	1	517.318	258.248	517.318
L19	1	1	537.746	537.746	537.746
L20	0.849	0.958	590.601	399.189	590.601
L21	0.787	0.942	622.550	378.043	622.550
L22	0.785	0.701	660.164	518.192	660.164
L23	1	1	808.369	808.369	808.369

7.3.3 Non-discretionary Variable (NDSC) Model

The NCN model bases on the formulae (7.3). The non-controllable constraints (7.4) and (7.5) demand that the non-controllable input/output variables on the left should be expressed exactly as *equality* by a nonnegative combination of the corresponding non-controllable variables in the data set on the right. This reflects situations like the following. When comparing the performance of supermarkets using the population of the district as the non-controllable variable, the right side population should be compared with the same size one as expressed by the left side of the equation.

However, if other situations (constraints) are preferred, i.e., *greater than or equal* (\geq) constraints in (7.4) and *less than or equal* (\leq) constraints in (7.5), the NDSC (non-discretionary variable) model can be utilized. This is the model introduced in (7.1), i.e.,

$$\begin{aligned} \min \quad & \theta - \varepsilon \left(\sum_{i \in D} s_i^- + \sum_{r=1}^s s_r^+ \right) \\ \text{subject to} \quad & \theta x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in D \\ & x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^-, \quad i \in ND \\ & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+, \quad r = 1, \dots, s. \end{aligned} \tag{7.6}$$

where all variables (except θ) are constrained to be nonnegative.

Although the non-discretionary (ND) variables do not enter the objective in (7.6) they, nevertheless, affect the efficiency score. Turning to the dual, as represented in (7.7), below, we can assign these effects a numerical value. This can be seen by writing the dual to (7.6) as follows.

$$\begin{aligned} \max_{\mathbf{v}, \mathbf{u}} \quad & \sum_{r=1}^s u_r y_{ro} - \sum_{i \in ND} v_i x_{io} \\ \text{subject to} \quad & \sum_{r=1}^s u_r y_{rj} - \sum_{i \in ND} v_i x_{ij} - \sum_{i \in D} v_i x_{ij} \leq 0, \quad j = 1, \dots, n \\ & \sum_{i=1}^m v_i x_{io} = 1 \\ & v_i \geq \varepsilon, \quad i \in D \\ & v_i \geq 0, \quad i \in ND \\ & u_r \geq \varepsilon, \quad r = 1, \dots, s \end{aligned} \tag{7.7}$$

Notice now that the objective of the dual in (7.7) has a property that is opposite to that of (7.6). Only the non-discretionary input multipliers enter into the objective of (7.7), i.e., the multiplier model, whereas only the discretionary (D) input variables enter into the objective of (7.6), i.e., the envelopment model. Thus, a value of the optimal multiplier v_i^* provides the *rate* at which the optimal value of (7.6) is reduced by increasing x_{io} for this member of the ND and $v_i^* x_{io}$ is the *amount* by which the optimal value of (7.6) is reduced by this member of the set ND when the corresponding constraint is critical in the envelopment model. On the other hand if this constraint is loose, so that positive slack is present for this constraint, the optimal value of the objective in (7.6) is not affected. Because this non-discretionary factor is present in more than needed amounts the optimal $v_i^* = 0$.

We illustrate the latter property by forming the dual to the preceding problem in Section 7.2 used to evaluate DMU P_5 in Figure 7.1,

$$\begin{aligned} & \max u - 2v_2 \\ & \text{subject to} \\ & 1 = 2v_1 \\ & 0 \geq u - v_1 - 2v_2 \\ & 0 \geq u - v_1 - \frac{3}{2}v_2 \\ & 0 \geq u - 2v_1 - v_2 \\ & 0 \geq u - 3v_1 - v_2 \\ & 0 \geq u - 2v_1 - 2v_2 \\ & v_1, u \geq \varepsilon, v_2 \geq 0. \end{aligned}$$

This has the following optimum solution,

$$u^* = v_1^* = 1/2$$

with all other variables zero, including $v_2^* = 0$. Hence the presence of the non-discretionary variable represented by x_{25} has no effect in the performance evaluation.

In the above example, no harm is experienced by $x_{25} = 2$ being in excess of what is needed since the slack variable makes it unnecessary to use all of what is available. However, there are cases in which an excessive amount of an input may cause a reduction in performance. This can occur with a discretionary variable as in the case of congestion where too much of a discretionary input is used in excessive amounts that cause a reduction in output. See Deng (2003).³ It can also occur with non-discretionary variables. An excessive amount of rain that damages farm crops is an example. The latter kind of excess will show up in the objective of the multiplier model when the non-discretionary variables are evaluated explicitly while the former will show up in the dual (envelopment model). See Problem 7.2, below.

7.3.4 Bounded Variable (BND) Model

There are several extensions of the these models —NCN and NDSC— among which we will introduce models with upper/lower bounds constraints. For example, when we evaluate the efficiency of baseball stadiums and take the number of spectators as an output variable, this value cannot exceed the maximum capacity of the stadium for each DMU. Thus, the maximum capacity should be considered as an upper bound for the number of spectators.

In order to cope with these situations, we will relax the non-controllable constraints (7.4) and (7.5) into constraints with upper/lower bounds as follows:

$$l_o^{N_x} \leq X^N \lambda \leq u_o^{N_x} \quad (7.8)$$

$$l_o^{N_y} \leq Y^N \lambda \leq u_o^{N_y} \quad (7.9)$$

where $(l_o^{N_x}, u_o^{N_x})$ and $(l_o^{N_y}, u_o^{N_y})$ are vectors of the lower and upper bounds to the non-discretionary inputs and outputs of DMU_{*o*}, respectively. x_o^N and y_o^N are not explicitly included in the formulae, because it is assumed that they lie in between the two bounds. The extensions including bounded non-discretionary variables are as follows:

(a) Input Oriented Bounded Variable Model

$$(BND_o) \quad \begin{aligned} & \min \quad \theta \\ & \text{subject to} \quad \theta x_o^C \geq X^C \lambda \\ & \quad y_o^C \leq Y^C \lambda \\ & \quad l_o^{N_x} \leq X^N \lambda \leq u_o^{N_x} \\ & \quad l_o^{N_y} \leq Y^N \lambda \leq u_o^{N_y} \\ & \quad L \leq e\lambda \leq U \\ & \quad \lambda \geq 0. \end{aligned} \quad (7.10)$$

(b) Output Oriented Bounded Variable Model

$$(BNDO_o) \quad \begin{aligned} & \max \quad \eta \\ & \text{subject to} \quad x_o^C \geq X^C \lambda \\ & \quad \eta y_o^C \leq Y^C \lambda \\ & \quad l_o^{N_x} \leq X^N \lambda \leq u_o^{N_x} \\ & \quad l_o^{N_y} \leq Y^N \lambda \leq u_o^{N_y} \\ & \quad L \leq e\lambda \leq U \\ & \quad \lambda \geq 0. \end{aligned} \quad (7.11)$$

7.3.5 An Example of the Bounded Variable Model

As an example of the bounded variable model, we introduce the evaluation of the relative efficiency of the twelve Japanese professional baseball teams in the year 1993. We employ 2 inputs and 2 outputs for this evaluation as follows:

Input 1 : Average annual salary of managers, including those of coaching staff.

Input 2 : Average annual salary of players: the top ranked 9 fielders and 6 pitchers.

Output 1 : “Team power” as explained below.

Output 2 : “Attendance” as explained below.

“Team power” is measured as a function of the percentage of victories per year, the batting averages of fielders, the number of home runs, the number of stolen bases and the defense rating of pitchers. We utilized results from a “principal component” analysis to define the team power of each team. Actually, the contribution ratio of the first principal component was 53% and was judged to be interpretable as “Team power.” The second component seems to represent “Mobile power” and its proportion is 23%. For this study we employed “Team power” as a representative index of the total power of the team. “Attendance” is measured as the ratio of the total annual attendance vs. the annual maximum capacity of the team’s home stadium and expressed as a percentage. Hence this number cannot exceed 100%. Table 7.3 exhibits these data for 12 teams. Salaries of managers and players, as inputs, are in tens of thousands of Japanese yen (ten thousand Japanese yen \approx 83 U.S. dollars in 1993).

Table 7.3. Data of 12 Japanese Baseball Teams in 1993

Team	INPUT		OUTPUT	
	Manager	Player	Team power	Attendance
Swallows	5250	6183	21.72	80.87
Dragons	2250	5733	21.02	93.76
Giants	6375	8502	18.44	100.00
Tigers	3125	4780	18.49	76.53
Bay Stars	3500	4042	18.05	79.12
Carp	3125	5623	18.55	51.95
Lions	5500	10180	21.25	56.37
Fighters	3625	5362	20.74	57.44
Blue Wave	2715	4405	20.39	58.78
Buffalos	3175	6193	20.59	53.00
Marines	2263	5013	17.78	43.47
Hawks	3875	3945	16.80	82.78
Average	3732	5830	19.49	69.51

The purpose of this study is to evaluate the relative efficiency of these 12 teams using salaries as inputs and seasonal records as outputs. Salary control,

especially reduction in the average salary, is very difficult for the owner of a team, so we employed the output-oriented models.

Using an output-oriented CCR model we obtained the results exhibited in Table 7.4 under the heading CCR, where the “Projection” of “Attendance” to efficient frontiers is also recorded in the next to last column. Full efficiency was attained by the Dragons, Bay Stars, Blue Wave and Hawks within this model. These teams had good outputs relative to comparatively low salaries. In contrast, the well paid Giants and Lions ranked as the lowest two. As can be seen, the CCR-Projections of Swallows, Giants and Lions exceed 100%. This, however, is impossible because of the capacity limit. Thus, the efficiency scores obtained under this assumption are not always matched to their “Attendance” situation.

Table 7.4. Projection of Attendance by CCR and Bounded Models

Team	CCR Score	Bounded Score	Attendance	CCR Projection	Bounded Projection
Swallows	0.7744	0.7589	80.87	104.40	82.51
Dragons	1	1	93.76	93.76	93.76
Giants	0.6139	0.5088	100.00	162.90	100.00
Tigers	0.9204	0.8829	76.53	83.15	76.53
Bay Stars	1	1	79.12	79.12	79.12
Carp	0.7552	0.7552	51.95	79.61	79.61
Lions	0.4852	0.5474	56.37	146.25	100.00
Fighters	0.8356	0.8356	57.44	71.55	71.55
Blue Wave	1	1	58.78	58.78	58.78
Buffalos	0.7948	0.7948	53.00	91.27	91.27
Marines	0.9048	0.9048	43.47	78.01	78.01
Hawks	1	1	82.78	82.78	82.78

Now we treat the output “Attendance” as a bounded variable with an upper bound of 100% and use the status quo for every team as a lower bound. In this model, “Team power” is the only variable connected directly to the efficiency evaluation. “Attendance” is treated as a second output which constrains the feasible region and thus indirectly influences the objective value. “Team power” is connected to attendance, however, because when the team power increases then “Attendance” also grows. The results are exhibited in Table 7.4 under the heading “Bounded Projection.” As can be seen, all projections listed under this heading remain within the prescribed bounds. The efficiency scores evaluated under these bounds as listed under “Bounded Score” are (a) changed only slightly from those in the CCR case, (b) none of the previously identified efficient teams lost this status, and (c) even some of the teams with Attendance reduced by the 100% bound increased their efficiency score. See the increase

from 0.4852 to 0.5474 for the Lions.

Remark : It might be argued that the value under CCR projection provides information on where added capacity is needed. Information of this kind is also available from the dual (multiplier) values when the constraints associated with these bounds are critical. Notice, for instance, that these constraints are critical for the Giants and Lions, which are both at 100% under “bounded projection,” but this is not true for the Swallows even though this team’s CCR projection exceeded 100%.

7.4 DEA WITH CATEGORICAL DMUS

There are other managerial situations over which managers of particular organizations do not have total control. For instance, in evaluating the performance of a branch store of a supermarket, it is necessary to consider the sales environment of the store, including whether it has severe competition, is in a normal business situation, or in a relatively advantageous one. If we evaluate the efficiency of the above supermarkets as “scratch” players, the evaluation would be unfair to the stores in the highly competitive situation and would be too indulgent to the stores in the advantageous one. Hence we need to provide something in the way of a “handicap” when evaluating them.

A hierarchical category is suitable for handling such situations. As for the supermarket example, we classify stores facing severe competition as *category 1*, in a normal situation as *category 2* and in an advantageous one as *category 3*. Then we evaluate stores in category 1 only within the group, stores in category 2 with reference to stores in category 1 and 2 and stores in category 3 within all stores in the model. Thus, we can evaluate stores under operating handicaps which take into account their particular environments, and we also use this information to evaluate stores in higher categories.

There are two kinds of categorical variables. One is *non-controllable* by decision makers, as pointed out above. The other is under the control of decision makers. We will consider the non-controllable type first.

7.4.1 An Example of a Hierarchical Category

The public libraries in Tokyo (in Table 7.1, above) can be classified into three categories as follows. Category 1 consists of libraries in the business district of central Tokyo, category 2 in the shopping area around the business district, and category 3 in the residential area on the outskirts. (See the column headed by “Category” in Table 7.5.) Libraries in category 1 are in a severe situation for library utilization, since residents around them are few compared with libraries in the other two categories. Category 3 is in the most advantageous situation. So we evaluate the efficiency of libraries in category 1 only within the category while libraries in category 2 are evaluated with reference to categories 1 and 2 and libraries in category 3 are evaluated with reference to all libraries. We utilized the data in Table 7.1 excluding “population” from inputs. Thus three

Table 7.5. Categorization of Libraries

Library	Category	Categorization		CCR	
		Score	Reference	Score	Reference
L1	1	0.377	L6	0.226	L23
L2	1	0.879	L5 L6	0.638	L23
L3	1	0.936	L4 L6	0.54	L23
L4	1	1	L4	0.593	L23
L5	1	1	L5	0.911	L17 L23
L6	1	1	L6	0.745	L23
L7	2	0.743	L5 L9	0.65	L19 L23
L8	2	0.648	L5 L15	0.539	L17
L9	2	1	L9	0.907	L19 L23
L10	2	0.815	L9 L15	0.705	L19 L23
L11	2	0.646	L5 L9 L15	0.539	L17 L23
L12	2	0.835	L9	0.719	L19 L23
L13	2	0.794	L9 L15	0.657	L17 L23
L14	2	0.835	L9	0.715	L23
L15	2	1	L15	0.844	L17 L23
L16	2	0.687	L9 L15	0.582	L23
L17	3	1	L17	1	L17
L18	3	0.787	L19 L23	0.787	L19 L23
L19	3	1	L19	1	L19
L20	3	0.849	L19 L23	0.849	L19 L23
L21	3	0.787	L23	0.787	L23
L22	3	0.681	L23	0.681	L23
L23	3	1	L23	1	L23

inputs (area, books, staff) and two outputs (registered, borrowed) are evaluated under categorical conditions.

The results of computations based on this categorization are given in Table 7.5. It can be seen that the reference set for all DMUs in category 1 consists of only category 1 DMUs. In the reference set for DMUs in category 2, however, L5 is included from category 1, and L23 appears only in the reference set of category 3. This is in contrast to the CCR model applied to all libraries altogether in which case L17, L19 and L23 have full efficiency score and appear as reference DMUs used to evaluate other libraries. See the column headed "CCR" in Table 7.5. These three libraries belong to category 3, i.e. the area which is most favorable to library utilization.

7.4.2 Solution to the Categorical Model

The categorical model can be incorporated into any DEA model, such as CCR, BCC, IRS, DRS and GRS which are discussed in Section 5.7 of Chapter 5. In

designing LP computations for the categorical model, we should notice that the DMUs in the upper categories cannot be chosen as basic variables for DMUs in a lower category.

7.4.3 Extension of the Categorical Model

In the preceding subsections, we dealt with the non-controllable category case, in which decision makers cannot choose one level from a category at will. However, in some cases, the category is under the control of decision makers. For example, suppose that the DMUs are a set of shops with three levels of service, i.e., *poor*, *average* and *good*. A shop owner in the *poor* service category has the option to remain *poor* or upgrade to *average* or *good*. An *average* shop can move within the *average* level or further up to the *good* level.

Suppose that each shop (DMU) can be rated as having one of L different categorical classes, where the categories range from category 1 (representing the lowest service orientation) to category L (the highest). The problem here is then to find, for each DMU, the DEA projected point in the same or higher category levels.

In the algorithm below, we consider the case for DMU_o , which is currently at level l ($1 \leq l \leq L$) and try to find the reference set and the DEA projected points on the frontier with levels in the same category or higher. As for the DEA model employed, we can choose any model, e.g. CCR, BCC, Additive, etc.

[Algorithm] (Controllable Categorical Level)

For $h = l, l + 1, \dots, L$, repeat the following steps:

Step 1.

Organize a set of DMUs composed of level h or *higher* and the DMU_o to be evaluated.

Evaluate the efficiency of DMU_o with respect to this group via the DEA model chosen. Go to Step 2.

Step 2.

- (i) If DMU_o is found to be efficient, go to Step 3.
- (ii) If DMU_o is inefficient, then record its reference set and reference (projected) point on the frontier.
If $h = L$, go to Step 3. Otherwise, replace h by $h + 1$ and go back to Step 1.

Step 3.

Examine the reference set, reference point, and category level, obtained from Step 2, and choose the most appropriate point and category level for DMU_o .

One of the characteristics of this algorithm is that it allows DMUs at different levels to be used as a candidate reference point. For example, suppose a *poor* DMU_1 has its reference set composed of an *average* DMU_2 and a *good* DMU_3 with weights 0.75 and 0.25. A categorical service level of $0.75 \times \text{average} + 0.25 \times \text{good}$ is assumed for the projected point, i.e. it has a quality level close to *average* and slightly upgraded to *good*.

[Example]

Table 7.6 exhibits nine single input and single output DMUs, each having either a *poor*, *average* or *good* category level. Figure 7.2 shows them graphically, where the symbols \bullet , \blacktriangle and \blacksquare correspond to *poor*, *average* and *good*, respectively. We applied this algorithm using the BCC model and obtained the results shown in the far right column of Table 7.6, where the number in parenthesis designates the λ -value assigned to the referent DMU.

Table 7.6. Nine DMUs with Three Category Levels

DMU	Input	Output	Category	Reference set
A	3	1	poor	A(1)
B	7	7	poor	D(.6), E(.4)
C	12	6	poor	D(.8), E(.2) G(1)
D	4	5	average	D(1)
E	6	10	average	E(1)
F	11	11	average	E(.667), H(.333) G(1)
G	8	11	good	G(1)
H	9	13	good	H(1)
I	13	15	good	I(1)

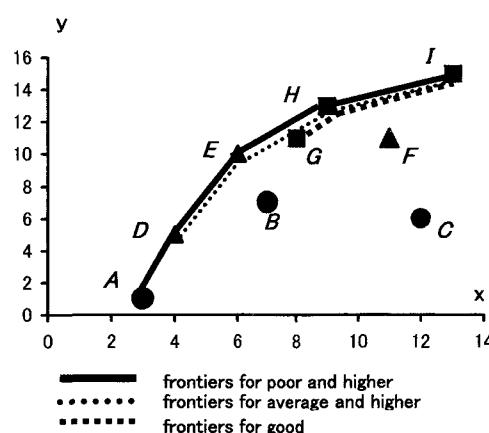


Figure 7.2. DMUs with Controllable Category Levels

DMU *A* (*poor*) is judged to be efficient, even compared with DMUs in the same or higher category levels and it is reasonable for *A* to remain in this category. DMU *B* (*poor*) is enveloped by *D* (*average*) with $\lambda_D = 0.6$ and *E* (*average*) with $\lambda_E = 0.4$ and is suitable for upgrading to *average*. DMU *C* (*poor*) has two sets of reference points, one composed of *D* (*average*) with $\lambda_D = 0.8$ and *E* (*average*) with $\lambda_E = 0.2$ and the other *G* (*good*) with $\lambda_G = 1$. So, two possibilities are available for DMU *C* to upgrade its level. Similarly, DMU *F* has two possibilities and this kind of information can be taken into account by higher management, for instance, in assigning bonuses, etc., based on actual or attained performance.

7.5 COMPARISONS OF EFFICIENCY BETWEEN DIFFERENT SYSTEMS

The DEA models assume that the production possibility set P is convex and, in fact, if two activities $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2)$ belong to P , then every point on the line segment connecting these two points belongs to P .⁴ However, there are situations where this assumption is not valid. For example, an activity $(\mathbf{x}_1, \mathbf{y}_1)$ uses one kind of instrument, while an activity $(\mathbf{x}_2, \mathbf{y}_2)$ adopts another, so we cannot reasonably assume any activity exists in between them.

7.5.1 Formulation

To see how this problem can be treated suppose the DMUs under consideration belong exclusively to one of two systems i.e. Systems *A* and *B*. (Although we deal with two systems, the discussions below can be easily extended to more general cases). We divide input X into X_A and X_B and output Y into Y_A and Y_B . The convexity assumption holds *within* the same system but does not hold *between* the two systems. The production possibility set $\{(\mathbf{x}, \mathbf{y})\}$ is then assumed to satisfy the following constraints:

$$\begin{aligned} \mathbf{x} &\geq X_A \boldsymbol{\lambda}_A + X_B \boldsymbol{\lambda}_B \\ \mathbf{y} &\leq Y_A \boldsymbol{\lambda}_A + Y_B \boldsymbol{\lambda}_B \\ Lz_A &\leq e\boldsymbol{\lambda}_A \leq Uz_A \\ Lz_B &\leq e\boldsymbol{\lambda}_B \leq Uz_B \\ z_A + z_B &= 1 \\ \boldsymbol{\lambda}_A &\geq \mathbf{0}, \boldsymbol{\lambda}_B \geq \mathbf{0} \\ z_A, z_B &= 0 \text{ or } 1. \end{aligned} \tag{7.12}$$

That is, the problem is found to be an integer program.

In this situation, the efficiency of DMU $(\mathbf{x}_o, \mathbf{y}_o)$ can be evaluated by the following mixed integer LP problem with z_A, z_B as binary variables that assume only the values 0 and 1,

$$\begin{array}{ll} \min & \theta \\ \text{subject to} & \theta \mathbf{x}_o \geq X_A \boldsymbol{\lambda}_A + X_B \boldsymbol{\lambda}_B \end{array} \tag{7.13}$$

$$\begin{aligned}
y_o &\leq Y_A \lambda_A + Y_B \lambda_B \\
Lz_A &\leq e\lambda_A \leq Uz_A \\
Lz_B &\leq e\lambda_B \leq Uz_B \\
z_A + z_B &= 1 \\
\lambda_A &\geq \mathbf{0}, \quad \lambda_B \geq \mathbf{0} \\
z_A, z_B &= \{0, 1\}.
\end{aligned}$$

From the results secured, we can evaluate not only the efficiency of each DMU but we can also compare the two systems by observing the efficiency of DMUs in each system. For this purpose, the nonparametric rank-sum test statistics described in the next section will be useful.

7.5.2 Computation of Efficiency

As is shown below, we can solve (7.13) by enumeration rather than by using a mixed integer 0-1 program,

1. Set $z_A = 1, z_B = 0$ and solve the LP problem above. That is, we evaluate the efficiency of DMU (x_o, y_o) , based on System A. Let the optimal objective value be θ_A . If the corresponding LP problem is infeasible, we define $\theta_A = \infty$.
2. Set $z_A = 0, z_B = 1$ and solve the LP problem above. Let the optimal objective value be θ_B , which is infinity if the corresponding LP problem is infeasible.
3. We obtain the efficiency of DMU (x_o, y_o) by

$$\theta_o^* = \min\{\theta_A, \theta_B\}.$$

7.5.3 Illustration of a One Input and Two Output Scenario

Figure 7.3 exhibits DMUs in two sales systems, A and B , with one input and two outputs. Output 1 is the number of customers per salesman and output 2 is profit per salesman. In this case, if we neglected the distinction between the systems, the efficient frontier would be the line segments connecting A1, A2, B7 and B10. For reasons described above, however, the efficient frontier is represented by the bold lines connecting A1, A2, A6, P, B7 and B10 which no

longer span convex production possibility sets like those we have employed in many of our previous examples.

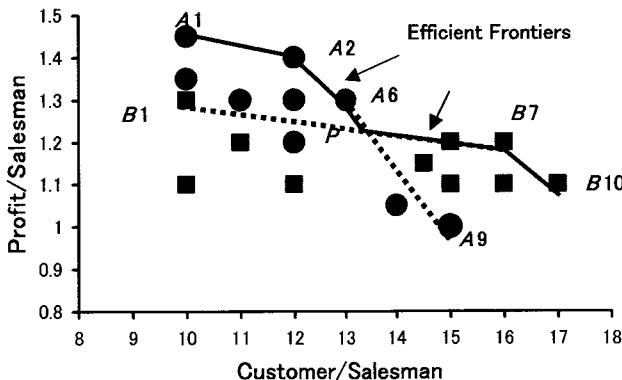


Figure 7.3. Comparisons between Stores using Two Systems

Table 7.7 shows the data set of this example under the heading “Data,” while the efficiency of stores obtained by using the algorithm and the CCR model, i.e. $L = 0, U = \infty$, are listed under the heading “Comparisons” in the same table. Since the efficient frontiers of System B are spanned by line segments connecting B_1, B_7 and B_{10} , it can be seen that DMUs in System A which are above these frontiers have the efficiency score θ_B greater than one as displayed in the table. Similarly, the efficient frontiers of System A are spanned by A_1, A_2, A_6 and A_9 and hence DMUs in B which are above the frontiers have the efficiency value θ_A greater than one. A comparison of the two systems is shown in Table 7.8 where we notice that it is possible that the reference set includes inefficient DMUs (e.g. B_1).

7.6 RANK-SUM STATISTICS AND DEA

As an instance of DEA applications of these ideas, comparisons of efficiencies between two types of DMUs, e.g. department stores vs. supermarkets, or electric power industries in the U.S.A vs. those in Japan, have constituted important research subjects. It is often necessary to test statistically the difference between two groups in terms of efficiency. Do differences occur by chance or are they statistically significant? This section deals with such statistical issues. Since the theoretical distribution of the efficiency score in DEA is usually unknown, we are forced to deal with nonparametric statistics for which the distribution of the DEA scores are statistically independent. For this purpose, the rank-sum-test developed by Wilcoxon-Mann-Whitney may be used to identify whether the differences between two groups are significant.

Table 7.7. Comparisons of Stores in Two Systems

Store	Data			Comparisons			
	In.	Out1	Out2	θ_A	θ_B	θ^*	Ref. Set
A1	1	10	1.45	1	1.115	1	A1
A2	1	12	1.4	1	1.091	1	A2
A3	1	10	1.35	0.941	1.038	0.941	A1 A2
A4	1	12	1.3	0.926	1.011	0.926	A1 A2
A5	1	12	1.3	0.962	1.023	0.962	A2 A6
A6	1	13	1.3	1	1.034	1	A6
A7	1	12	1.2	0.923	0.955	0.923	A6
A8	1	14	1.05	0.969	0.875	0.875	B7
A9	1	15	1.0	1	0.893	0.893	B7 B10
B1	1	10	1.3	0.912	1	0.912	A1 A2
B2	1	11	1.2	0.885	0.943	0.885	A2 A6
B3	1	10	1.1	0.808	0.864	0.808	A2 A6
B4	1	12	1.1	0.892	0.886	0.886	B1 B7
B5	1	14.5	1.15	1.023	0.949	0.949	B1 B7
B6	1	15	1.2	1.062	0.989	0.989	B1 B7
B7	1	16	1.2	1.108	1	1	B7
B8	1	15	1.1	1.031	0.929	0.929	B7 B10
B9	1	16	1.1	1.077	0.964	0.964	B7 B10
B10	1	17	1.1	1.133	1	1	B10

Table 7.8. Comparisons of Two Systems

System	System A	System B
No. Eff. Stores	3	2
Average Eff.	0.947	0.932
Frequency of Reference to Other System	6	3

7.6.1 Rank-Sum-Test (Wilcoxon-Mann-Whitney)

This method is one of the nonparametric statistical tests based on the ranking of data. Given statistically independent data belonging to two groups, this test serves to test whether the hypothesis that the two groups belong to the same population or whether they differ significantly.

Let the data in two groups be represented by $A = \{a_1, a_2, \dots, a_m\}$ and $B = \{b_1, b_2, \dots, b_n\}$. We then merge A and B to arrive at a sequence C in

which the data are arranged in descending order. An example is as follows,

$$A = \{0.42, 0.51, 0.45, 0.82, 0.9, 1, 0.72, 0.92, 0.65, 0.87\} \quad (m = 10),$$

$$B = \{0.12, 0.32, 0.18, 0.42, 0.56, 0.9, 0.7, 0.22, 0.44, 0.02, 1\} \quad (n = 11).$$

We then merge these sequences into a new sequence, C , with length $m + n = 10 + 11 = 21$:

$$C = \{\underline{1}, 1, \underline{0.92}, \underline{0.9}, 0.9, \underline{0.87}, \underline{0.82}, \underline{0.72}, 0.7, \underline{0.65}, 0.56, \underline{0.51}, \underline{0.45}, 0.44, \\ \underline{0.42}, 0.42, 0.32, 0.22, 0.18, 0.12, 0.02\}$$

in which the underlined numbers belong to A .

Then we rank C from 1 to $N (= m + n)$. If there is a tie, we use the midrank for the tied observation.

The above example yields the ranking below,

$$R = \{\underline{1.5}, 1.5, \underline{3}, \underline{4.5}, 4.5, \underline{6}, \underline{7}, \underline{8}, 9, \underline{10}, 11, \underline{12}, \underline{13}, 14, \underline{15.5}, 15.5, 17, 18, 19, 20, 21\}.$$

For example, the top two numbers in the sequence C have the same value 1 so they are ranked as $(1+2)/2=1.5$ which is their midrank in R and the 4th and 5th in the sequence C have a tie, 0.9, so their midrank in R is $(4+5)/2=4.5$.

Next, we sum the ranking of the A data indicated by the underlined numbers,

$$S = 1.5 + 3 + 4.5 + 6 + 7 + 8 + 10 + 12 + 13 + 15.5 = 80.5.$$

Then statistic, S , follows an approximately normal distribution with mean $m(m + n + 1)/2$ and variance $mn(m + n + 1)/12$ for $m, n \geq 10$. By normalizing S , we have:

$$T = \frac{S - m(m + n + 1)/2}{\sqrt{mn(m + n + 1)/12}}. \quad (7.14)$$

T has an approximately standard normal distribution. Using T we can check the null hypothesis that the two groups have the same population at a level of significance α . We will reject the hypothesis if $T \leq -T_{\alpha/2}$ or $T \geq T_{\alpha/2}$, where $T_{\alpha/2}$ corresponds to the upper $\alpha/2$ percentile of the standard normal distribution. In the above example, we have $T = -2.0773$. If we choose $\alpha = 0.05$ (5%), then it holds that $T_{0.025} = 1.96$. Since $T = -2.0773 < -1.96 = -T_{0.025}$, we reject the null hypothesis at the significance level 5%. This test, which is attributed to Wilcoxon, is essentially equivalent to the Mann-Whitney test.

7.6.2 Use of the Test for Comparing the DEA Scores of Two Groups

A simple and straightforward application of the test mentioned above is to use it for identifying the difference between DEA scores in two groups of DMUs. Let the two groups be A and B . We pool them into $C (= A \cup B)$. Then, we employ an appropriate DEA model to measure the efficiency of each DMU in C . The rank-sum-test can be applied to check the hypothesis that groups A and B have the same population of efficiency scores.

7.6.3 Use of the Test for Comparing the Efficient Frontiers of Two Groups

Sometimes, it is required to compare the efficient frontiers of group A with those of group B and to see whether there is a significant frontier shift between the two.

For this purpose, we first run DEA separately for the two groups (once for A and once for B) and obtain the efficient DMUs in each group. Although we can compare the two groups using these efficient DMUs within each group, the number of DMUs for evaluation will decrease to a small portion of the total, and we will be confronted with a degrees-of-freedom problem for comparisons. To avoid this difficulty, we project inefficient DMUs in the group A (B) onto their respective efficient frontiers to obtain a new (combined) efficient frontier consisting of the same number of DMUs as in the original group. Then, we pool the two groups of efficient DMUs and run DEA. The rank-sum-test will be applied to the DEA results.

This method, which will eliminate inefficiency within each group makes it possible to compare the performance of two types of organization when both are operating efficiently.

7.6.4 Bilateral Comparisons Using DEA

The preceding two methods use a *mixed* evaluation of DMUs in two groups, by which it is meant that each DMU is evaluated on both *inter* and *within* group bases. Thus, the reference set of a DMU in A may consist of DMUs in A (*within*), or B (*inter*). In contrast to this, there is a view that each DMU in A (B) should be evaluated with respect to DMUs in the opposite group B (A). This *inter* comparison will result in a sharper discrimination between two groups. We can formulate this idea in the following way for each DMU $a \in A$.

$$\min \quad \theta \quad (7.15)$$

$$\text{subject to} \quad \sum_{j \in B} x_j \lambda_j \leq \theta x_a \quad (7.16)$$

$$\sum_{j \in B} y_j \lambda_j \geq y_a \quad (7.17)$$

$$\lambda_j \geq 0 \quad (\forall j \in B). \quad (7.18)$$

In the two input (x_1, x_2) and single unitized output ($y = 1$) case, Figures 7.4 and 7.5 illustrate a use of this program. In Figure 7.4, the DMU $a \in A$ is enveloped by DMUs in B and the optimal θ^* is given by

$$\theta^* = \frac{\text{OQ}}{\text{OP}} < 1.$$

In Figure 7.5, the DMU a cannot be enveloped by DMUs in B and a should be expanded radially to Q to be enveloped. The optimal θ^* is given by

$$\theta^* = \frac{\text{OQ}}{\text{OP}} > 1.$$

Thus, the optimal θ^* is no longer bounded by 1. The region of θ^* is therefore

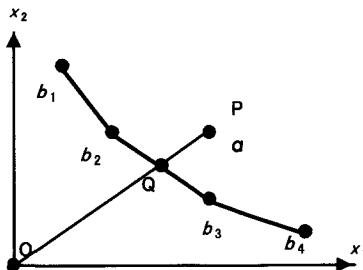


Figure 7.4. Bilateral Comparisons—Case 1

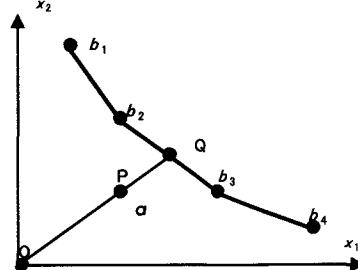


Figure 7.5. Bilateral Comparisons—Case 2

expanded upward and a sharper contrast between two groups can be expected than in the case of the *mixed* evaluation scheme. In order to make the LP feasible, we assume $X > O$ and $Y > O$. Bilateral comparisons of DMUs in B with respect to A follow in the same way. By virtue of its nonparametric character we can then apply the rank-sum-test to the thus obtained efficiency score θ^* , whether or not the groups A and B have the same distribution of efficiency values.

7.6.5 An Example of Bilateral Comparisons in DEA

We applied the bilateral comparisons to the library problem exhibited in Table 7.5 for testing the statistical difference in efficiency between the libraries in the business district (Category 1) and those in the residential area (Category 3). Thus Group A consists of six libraries L1-L6 and Group B consists of seven libraries L17-L23. We utilized numbers of books and staffs as inputs and numbers of registered residents and borrowed books as outputs (see Table 7.1). The results from these comparisons are exhibited in Table 7.9, where the efficiency scores against the counterpart libraries are listed along with the reference set and the ranking of the efficiency scores. The rank sum of Group A is $S = 63$ and the statistic calculated from (7.14) gives

$$T = \frac{63 - 6 \times (6 + 7 + 1)/2}{\sqrt{6 \times 7 \times (6 + 7 + 1)/12}} = 3.$$

Since, for $\alpha = 0.01$ (1%), the upper $\alpha/2$ percentile of the standard normal distribution is $T_{0.005} = 2.58$, we reject the null hypothesis that the efficiency scores of these two groups belong to the same distribution at the 1% significance level. That is, Group B outperforms Group A in efficiency.

Table 7.9. Example of Bilateral Comparisons

Group	DMU	Score	Reference set	Rank
Against B				
A	L1	0.226006153	L23	13
A	L2	0.637737534	L23	10
A	L3	0.540054844	L23	12
A	L4	0.593020886	L23	11
A	L5	0.911284882	L17 L23	8
A	L6	0.744964289	L23	9
Average		0.608844765	Rank Sum	63
Against A				
B	L17	1.283765162	L5 L6	5
B	L18	1.309557862	L6	4
B	L19	1.769104754	L6	1
B	L20	1.408073355	L6	3
B	L21	1.26630526	L5 L6	6
B	L22	1.000331642	L6	7
B	L23	1.612422038	L5 L6	2
Average		1.378508582	Rank Sum	28

7.6.6 Evaluating Efficiencies of Different Organization Forms

We turn to a study of the U.S. insurance industry and show how the approach described in Section 7.6.3 can be used to address a problem that has not been satisfactorily addressed by other methods. For this purpose we note that the U.S. insurance industry has long been dominated by two different forms of organization: “mutual companies” which are organized in the manner of co-operatives so that policyholders are also the owners of the company and (2) “stock companies” which are organized in the manner of private corporations with stockholders as owners and policy holders as customers. This has led to many studies directed to discovering whether one of these organization forms is superior (i.e., more efficient) than the other.

For instance, using data on various accounting ratios — such as rate of return on sales, etc., — a supposedly decisive study by Fama and Jensen (1983) concluded that each of these two forms was efficient — indeed, more efficient — for the niche it occupied.⁵ The presumption was that organizations compete for survival. The fact that both forms continue to survive therefore documents the claimed result.

In a more recent study, however, Brockett *et al.* (1998)⁶ argued that, first, Fama and Jensen confused observed performance (efficient or not) with the

potential capabilities of these two types of organization. Second, Brockett *et al.* argue that a use of ratios used separately on the assumption of independence of the resulting tests did not provide a satisfactory basis for evaluating the performance of institutions, such as insurance companies, which must have multiple interacting objectives if they are to meet the needs of an industry charged with a "public interest."⁷ Brockett *et al.* — turning to DEA — eliminated the need for *assuming* that technical efficiency is attained by every DMU — an assumption which underlies the "agency theory" approaches used by Fama and Jensen.

To satisfy the first and third of these study objectives, Brockett *et al.* followed the approaches described in the immediately preceding section of this chapter. Thus, in agreement with Section 7.6.3 they identified two efficient frontiers — one for mutual companies and another for stock companies. They then used the results from applying DEA to the data for each of these two groups to project each DMU onto its respective (mutual or stock) efficient frontiers. See the formulas in (3.28) and (3.29). This eliminated the need for *assuming* that efficiency had been attained and it also moved each DMU up to the boundary of the capabilities which the evidence showed was possible for its form of organization.

This done, Brockett *et al.* again utilized DEA to establish a new (combined) frontier against which each entity (=mutual or stock company) could be accorded a new (relative) efficiency rating. Note therefore that this rating showed how each entity could perform under the condition that it had already achieved the best performance that its form of organization allowed. Finally the Mann-Whitney test procedures were applied (a) to see if the rankings for two groups differed significantly and (b) to determine which group dominated when the hypothesis maintained under (a) was rejected.

This study covered other items such as the forms of sales and marketing systems used by these companies. However, we do not go into further detail and focus, instead, on the issue of relative efficiency of these two organization forms. Thus applying the approach we just described to 1989 data for $n_1=1114$ stock companies and $n_2=410$ mutual companies in the U.S. property-liability insurance industry,⁸ Brockett *et al.* concluded that the stock form was uniformly more efficient (and significantly so) in all of the dimensions studied. *Inter alia*, they documented this finding by noting that stock companies could merge and acquire (or be acquired by) other companies in ways that were not available to mutual companies in order to obtain access to resources, talents and markets. Recent activities in "de-mutualization" by some of the largest U.S. mutual companies tends to support the study results, especially when it is noted that no movements have occurred in the opposite direction. Hence we conclude with Brockett *et al.* that this use of DEA uncovered latent possibilities that were not detected in the approaches used by others, as well as Fama and Jensen.

7.7 SUMMARY OF CHAPTER 7

In this chapter we expanded the ability of DEA to deal with variables that are not under managerial control but nevertheless affect performances in ways that need to be taken into account when effecting evaluations. Non-discretionary and categorical variables represent two of the ways in which conditions beyond managerial control can be taken into account in a DEA analysis. Uses of upper or lower bounds constitute yet another approach and, of course, these approaches can be combined in a variety of ways. Finally, uses of Wilcoxon-Mann-Whitney statistics were introduced for testing results in a nonparametric manner when ranking can be employed.

Illustrative examples were supplied along with algorithms that can be used either separately or with the computer code DEA-Solver. We also showed how to extend DEA in order to deal with production possibility sets (there may be more than one) that are not convex. Finally we provided examples to show how new results can be secured when DEA is applied to such sets to test the efficiency of organization forms (and other types of activities) in ways that were not otherwise available.

7.8 NOTES AND SELECTED BIBLIOGRAPHY

Banker and Morey (1986a) developed the first model for evaluating DEA efficiency with exogenously fixed, inputs and outputs, with other models following as in Charnes *et al.* (1987), Adolphson *et al.* (1991)⁹ and Tone (1993).¹⁰ (See also Problems 7.9 and 7.10.)

Models treating categorical inputs and outputs were also introduced by Banker and Morey (1986b).¹¹ They formulated the controllable categorical variable problem within the framework of a mixed-integer LP model under the BCC model. Kamakura (1988)¹² pointed out shortcomings of their formulation and presented another mixed-integer model. Rousseau and Semple (1993)¹³ presented a new algorithm for this case, as an extension of Kamakura's work (1988). This method eliminated difficulties of computation that had accompanied earlier mixed integer models. It should be noted that the algorithms introduced in Sections 7.3 and 7.4 can be coupled with any DEA model, in contrast to earlier methods. (See Tone (1997).¹⁴)

Tone (1995)¹⁵ proposed the bounded variable models in Section 7.3. The materials in Section 7.5 (comparisons of efficiency between different systems) was introduced by Tone (1993).¹⁶

The use of DEA to effect bilateral comparisons were first explored in Charnes, Cooper and Rhodes (1981).¹⁷ The rank-sum test in Section 7.5.3 was developed by Brockett and Golany (1996).¹⁸ See also Brockett and Levine (1984)¹⁹ for a detailed development.

7.9 RELATED DEA-SOLVER MODELS FOR CHAPTER 7

NCN-I-C(V),NCN-O-C(V) These codes solve the input (output)-oriented non-controllable variable model expressed by (7.3) under a constant (vari-

able) returns-to-scale assumption. The sample data format can be found in Figure B.3 in Section B.5.3 of Appendix B, where non-controllable input or output are designated by (IN) or (ON)), respectively. The results of the computation are recorded in worksheets “Summary,” “Score,” (“Rank,”) “Projection,” “Graph1,” and “Graph2.”

NDSC-I(O)-C(V, or GRS) These codes solve the input (output)-oriented non-discretionary variable model under the constant (variable or general) returns-to-scale assumption that is described in (7.6). The input data format is the same with the NCN.

BND-I-C(V or GRS), BND-O-C(V or GRS) These codes solve the input (output)-oriented bounded variable model expressed by (7.10) or (7.11) under a constant (variable or general) returns-to-scale assumptions. See Figure B.4 in Section B.5.4 of Appendix B for reference, where (IB)Doc., (LB)Doc. and (UB)Doc. designate “Doc.(doctor)” as a bounded variable and its lower/upper bounds (LB)Doc. and (UB)Doc. Joint use of an input variable and a bounded variable is permitted. It is thus possible to register Doc. as (I) and at the same time as (IB). See Problems 7.9 and 7.10. The results of the computation are recorded in worksheets “Summary,” “Score,” “Rank,” “Projection,” “Graph1,” and “Graph2.”

CAT-I-C(V), CAT-O-C(V) These codes solve the input (output)-oriented hierarchical category model as introduced in Section 7.4.1. See Figure B.5 in Appendix B for an example. The main body of the data format is the same as for the CCR case, with an additional column showing the categorical level. Its heading is arbitrary, e.g. Category. The category number ranges 1, 2, …, L, where DMUs in Level 1 are in the most disadvantageous condition and they will be compared only among themselves. Level 2 is in a better position than 1, and the DMUs in Level 2 will be compared with reference to DMUs in Levels 1 and 2 and so on. It is recommended that the level numbers be continuously assigned starting from 1. The results of the computation are recorded in worksheets “Summary,” “Score,” “Rank,” “Projection,” “Graph1,” and “Graph2.” Also, statistics by category class can be found in “Summary.”

SYS-I-C(V), SYS-O-C(V) These codes deal with comparisons of efficiency between different systems, as introduced in Section 7.5. The data format is the same as for the SYS model, with the last column showing the system number each DMU belongs to. In contrast to the CAT model, this number is used only for discrimination purposes but it is recommended that the system numbers be continuously assigned starting from 1. The results of the computation are recorded in worksheets “Summary,” “Score,” “Rank,” “Projection,” “Graph1,” and “Graph2.” Also, statistics by system class can be found in “Summary.”

Bilateral This code solves the bilateral comparisons model expressed by (7.15)-(7.18) in Section 7.6.4. It tests the null hypothesis that the efficiency of

DMUs in the concerned two groups belongs to the same statistical distribution using nonparametric rank-sum statistics. The data format is the same as for the CAT model, where the only level numbers allowed to be assigned are 1 or 2. The results of this test are shown in “Score,” as well as in “Summary.” This model has four variations as follows:

- Bilateral-CCR-I, Bilateral-BCC-I
- Bilateral-SBM-C, Bilateral-SBM-V

7.10 PROBLEM SUPPLEMENT FOR CHAPTER 7

Problem 7.1

Part (1): Formulate the dual to (7.1) and discuss its uses in evaluating performances.

Suggested Response: The dual to (7.1) can be represented as follows:

$$\begin{aligned} \max_{\mu, \nu} \quad & \sum_{r=1}^s \mu_r y_{ro} - \sum_{i \in ND} \nu_i x_{io} \\ \text{subject to} \quad & \sum_{r=1}^s \mu_r y_{rj} - \sum_{i \in ND} \nu_i x_{ij} - \sum_{i \in D} \nu_i x_{ij} \leq 1 \quad (j = 1, \dots, n) \\ & \sum_{i \in D} \nu_i x_{io} = 1 \\ & \nu_i \geq \varepsilon, i \in D; \quad \nu_i \geq 0, i \in ND; \quad \mu_r \geq \varepsilon, r = 1, \dots, s. \end{aligned}$$

Whereas (7.1) evaluates the discretionary variables in the set D in its objective, the above model evaluates the non-discretionary variables in the set ND in its objective. Between the two all possibilities for the evaluation are covered and are related to each other by the duality theorem of linear programming. See A.4 in Appendix A. Thus, if a ND constraint is critical it will have zero slack in the objective of (7.1) and $\nu_i^* > 0$ in the above dual (multiplier) model to evaluate the limitation it imposes on performance. If non-zero slack appears in (7.1) then the corresponding dual variable $\nu_i^* = 0$ in the multiplier model so this non-zero slack evaluates the complete discretion this constraint allows to management in the primal (=envelopment) model.

Part (2): Illustrate your interpretation by formulating and interpreting the dual to the problem that was used to evaluate P_5 immediately following Figure 7.1.

Suggested Response: Reference to the formulation of the dual in Problem 7.1 gives the following as the multiplier model that is dual to the model that was used to evaluate P_5 in the discussion following Figure 7.1,

$$\max \mu - 2\nu_2$$

$$\begin{aligned}
& \text{subject to} \\
& 1 = 2\nu_1 \\
& 0 \geq \mu - \nu_1 - 2\nu_2 \\
& 0 \geq \mu - \nu_1 - \frac{3}{2}\nu_2 \\
& 0 \geq \mu - 2\nu_1 - \nu_2 \\
& 0 \geq \mu - 3\nu_1 - \nu_2 \\
& 0 \geq \mu - 2\nu_1 - 2\nu_2 \\
& \mu, \nu_1 \geq \varepsilon; \nu_2 \geq 0.
\end{aligned}$$

The solution to this problem is $\mu^* = \nu_1^* = \frac{1}{2}$ with all other variables zero. This solution is confirmed by noting that $\mu^* = \theta^* = \frac{1}{2}$ as required by the duality theorem. Note, therefore, that $\nu_2^* = 0$ shows that this constraint did not enter into the evaluation of the performance of the DMU associated with P_5 .

Problem 7.2

Part (1): Modify the problem following Figure 7.1 to reflect the fact too much rainfall damaged a farmer's crop. Then discuss how this affects the above dual (multiplier) model formulation and interpretation.

Suggested Response: This can be most easily expressed by replacing the inequality for the ND constraint (representing rainfall) by an equation in the primal problem used to evaluate P_5 . This affects the dual, as formulated above, by allowing the ν_i associated with this constraint to assume negative as well as positive values and thus eliminates the condition $\nu_i \geq \varepsilon > 0$ for this variable.

Part (2): Apply the above modification to the preceding illustrative example.

Suggested Response: Replacing the inequality by an equation in the primal problem eliminates the alternate optimum possibility which had $s_2^- > 0$ in the solutions to (7.1). The uniquely optimal solution to this primal is then $\theta^* = \frac{1}{2}$, $\lambda_1^* = 1$ and all other variables zero. This replacement by an equation in the primal results in a solution to the dual with $\nu_1^* = \frac{1}{2}$, $\nu_2^* = -\frac{1}{4}$ and all other variables zero to yield $-2\nu_2^* = \theta^* = \frac{1}{2}$, as required by the duality theorem.

Note that this value of $\nu_2^* = -\frac{1}{4}$ means that performance could have been improved at the rate $\nu_2^* = -\frac{1}{4}$ if $x_{2o} = 2$ had been lowered in the primal. For instance, if $x_{2o} = 2$ were replaced by $\hat{x}_{2o} = 1$ the point would coincide with P_3 , which is efficient, as shown in Figure 7.1, with the uniquely optimal solution $\theta^* = 1$, $\lambda_3^* = 1$ and all other variables zero. Hence the entire inefficiency is caused by the excess rainfall.

Remark: The usual techniques used in the linear programming literature for exploring such changes are not applicable in DEA because the changes occur on both sides of the inequalities. See A. Charnes, W.W. Cooper, A. Lewin

and R. Morey (1985) "Sensitivity and Stability Analysis in DEA," Annals of Operations Research 2, pp.139-156.

Problem 7.3

Interpret (7.2) and discuss its use when (1) all $\beta_i, \gamma_r = 0$ and when (2) some or all of these parameters are assigned positive values.

Suggested Response : (1) If all $\beta_i, \gamma_r = 0$ then all inputs and outputs are non-discretionary. All slacks will be zero in the objective as well as the constraints. Hence no inefficiency is present and any solution is as good as any other solution because inefficiency must be associated with discretionary variables if it is to be correctable.

(2) If some or all parameters are assigned positive values then any nonzero slack which appears in the objective (a) incurs a penalty and (b) shows where corrections can be made. For instance, if all $\beta_i = 0$ and all $\gamma_r = \infty$ — so the latter collection of constraints can be omitted — then attention is focussed on the output improvements that are possible when all inputs are non-discretionary and have to be used in observed amounts.

Problem 7.4

Relate (7.3) to (7.2) and (7.1).

Suggested Response : If all $\beta_i, \gamma_r = 0$ in (7.2) then it can be simplified by omitting these constraints and replacing the corresponding inequality relations by equalities as in the sets indexed by N in (7.3). The remaining expressions in (7.3) can be brought into correspondence with (7.1) by (a) omitting the condition $L \leq e\lambda \leq U$, or, equivalently, setting $L = 0$ and $U = \infty$ and (b) omitting the slacks in the objective of (7.1).

Problem 7.5

Background : Recall from Section 7.2 in this chapter that an "active" vector is one which enters into a solution with a positive coefficient. We can express this mathematically as follows,

$$\begin{aligned}\theta_o^* x_{io} &= \sum_{j=1, j \neq k}^n x_{ij} \lambda_j^* + x_{ik} \lambda_k^* + s_i^{-*}, \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1, j \neq k}^n y_{rj} \lambda_j^* + y_{rk} \lambda_k^* - s_r^{+*}, \quad r = 1, \dots, s\end{aligned}\quad (7.19)$$

with all variables nonnegative. Here we have simplified (7.1) by assuming that non-discretionary inputs or outputs are not to be considered. Then the vector with components (x_{ik}, y_{rk}) is "active" if $\lambda_k^* > 0$.

We are using "*" to represent an optimum value in the following theorem.

Theorem 7.1 An input-output vector with components (x_{ik}, y_{rk}) will not enter actively into an optimum evaluation for any DMU_o unless the DMU_k with which it is associated is DEA-efficient.

Assignment : Prove this theorem.

Suggested response : Although we have proved this theorem in the form of Theorem 3.3 using the multiplier model in Chapter 3, we will give another proof which uses the envelopment model and extends to BCC as well as CCR models. Assume that DMU_k is not efficient so that either or both of the conditions for efficiency as specified in Definition 3.2 must fail to be satisfied. Then consider, first, the case in which DMU_k 's failure to be efficient is identified with $0 < \theta_k^* < 1$. The solution used to evaluate DMU_k can be written

$$\begin{aligned}\theta_k^* x_{ik} &= \sum_{j=1}^n x_{ij} \hat{\lambda}_j^* + \hat{s}_i^{-*}, \quad i = 1, \dots, m \\ y_{rk} &= \sum_{j=1}^n y_{rj} \hat{\lambda}_j^* - \hat{s}_r^{+*}, \quad r = 1, \dots, s\end{aligned}\tag{7.20}$$

with all variables nonnegative and $0 < \theta_k^* < 1$.

Substitution in (7.19) gives

$$\begin{aligned}\theta_o^* x_{io} &= \sum_{j=1, j \neq k}^n x_{ij} \lambda_j^* + s_i^{-*} + \frac{\lambda_k^*}{\theta_k^*} \left[\sum_{j=1}^n x_{ij} \hat{\lambda}_j^* + \hat{s}_i^{-*} \right], \quad i = 1, \dots, m \\ y_{ro} &= \sum_{j=1, j \neq k}^n y_{rj} \lambda_j^* - s_r^{+*} + \lambda_k^* \left[\sum_{j=1}^n y_{rj} \hat{\lambda}_j^* - \hat{s}_r^{+*} \right], \quad r = 1, \dots, s\end{aligned}\tag{7.21}$$

By hypothesis $0 < \theta_k^* < 1$. Also λ_k^* and its coefficients (in square brackets) are positive. Hence

$$\begin{aligned}\sum_{j=1}^n x_{ij} \lambda_j^* + s_i^{-*} &= \theta_o^* x_{io} > \sum_{j=1, j \neq k}^n x_{ij} \lambda_j^* + s_i^{-*} + \lambda_k^* \left[\sum_{j=1}^n x_{ij} \hat{\lambda}_j^* + \hat{s}_i^{-*} \right], \quad (7.22) \\ i &= 1, \dots, m \\ y_{ro} &= \sum_{j=1, j \neq k}^n y_{rj} \lambda_j^* - s_r^{+*} + \lambda_k^* \left[\sum_{j=1}^n y_{rj} \hat{\lambda}_j^* - \hat{s}_r^{+*} \right]. \quad r = 1, \dots, s\end{aligned}$$

Collecting terms on the right a solution is obtained which satisfies all constraints. Because strict inequality holds in all of the first m constraints, the value of θ_o^* can be reduced while continuing to satisfy all constraints. Hence the assumption $\lambda_k^* > 0$ for $0 < \theta_k^* < 1$ is not consistent with the optimality assumed for (7.19).

Now assume $\theta_k^* = 1$ but some slacks are not zero. In this case (7.21) becomes

$$\theta_o^* x_{io} = \sum_{j=1, j \neq k}^n x_{ij} \lambda_j^* + s_i^{-*} + \lambda_k^* \left[\sum_{j=1}^n x_{ij} \hat{\lambda}_j^* + \hat{s}_i^{-*} \right], \quad i = 1, \dots, m \tag{7.23}$$

$$y_{ro} = \sum_{j=1, j \neq k}^n y_{rj} \lambda_j^* - s_r^{+*} + \lambda_k^* \left[\sum_{j=1}^n y_{rj} \widehat{\lambda}_j^* - \widehat{s}_r^{+*} \right]. \quad r = 1, \dots, s$$

Now if any \widehat{s}_i^{-*} or $\widehat{s}_r^{+*} \neq 0$ then $\lambda_k^* > 0$ would multiply these nonzero slacks to yield a new set with a greater sum than $\sum_{i=1}^m s_i^{-*} + \sum_{r=1}^s s_r^{+*}$. Therefore the latter sum could not be maximal as required for optimality in the objective for the second stage maximization of the CCR model.

Thus, if either $0 < \theta_k^* < 1$ or some slacks are not zero in (7.20) then $\lambda_k^* > 0$ cannot hold in (7.19). The same is true, *a fortiori*, if both $0 < \theta_k^* < 1$ and some slacks are not zero. The theorem is therefore proved. DMU_k cannot be active in the evaluation of any DMU_o . \square

Corollary 7.1 *The above theorem is also true for the class of additive models.*

The proof of this corollary follows from the sum of slacks being maximal, as in case 2 of the above proof.

Corollary 7.2 *The above theorem, as proved for the CCR model, also holds for the BCC model.*

Proof : Non-negativity of the $\lambda_j^*, \widehat{\lambda}_j$ as required for the CCR model satisfies one part of the convexity condition prescribed for the BCC model. To show that the other condition is also satisfied when the BCC model is used instead of the CCR model, we collect the variables in (7.23) to define the following new set

$$\begin{aligned}\lambda_j^{new} &= \lambda_j^* + \lambda_k^* \widehat{\lambda}_j^*, \quad j \neq k \\ \lambda_k^{new} &= \lambda_k^* \widehat{\lambda}_k^*.\end{aligned}$$

Therefore

$$\sum_{j \neq k} \lambda_j^{new} + \lambda_k^{new} = \sum_{j \neq k} \lambda_j^* + \lambda_k^* \left(\sum_{j \neq k} \widehat{\lambda}_j^* + \widehat{\lambda}_k^* \right).$$

We now show that this is equal to unity when

$$\sum_{j=1}^n \lambda_j^* = \sum_{j=1}^n \widehat{\lambda}_j^* = 1.$$

This result is immediate if λ_k^* is zero or unity. Therefore assume $0 < \lambda_k^* < 1$, in which case we have

$$\lambda_k^* \left(\sum_{j=1}^n \widehat{\lambda}_j^* \right) = \lambda_k^*$$

and

$$\sum_{j \neq k} \lambda_j^* = 1 - \lambda_k^*$$

so, summing, we have

$$\sum_{j \neq k} \lambda_j^* + \lambda_k^* \left(\sum_{j \neq k} \widehat{\lambda}_j^* + \widehat{\lambda}_k^* \right) = \sum_{j=1, j \neq k}^n \lambda_j^{new*} + \lambda_k^{new*} = 1,$$

as claimed. \square

Problem 7.6

Using the point $P_1 = (1, 2)$ in Figure 7.1 and the model (7.2), show that the above theorem and corollaries need not hold when non-discretionary variables are involved.

Suggested Response : If $x_2 = 2$ is the only non-discretionary variable, the application of (7.2) to the data in Figure 7.1 yields

$$\begin{aligned} \max \quad & s_1^- + s_2^- + s^+ \\ \text{subject to} \quad & \lambda_1 + \lambda_2 + 2\lambda_3 + 3\lambda_4 + 2\lambda_5 + s_1^- = 1 \\ & 2\lambda_1 + 3/2\lambda_2 + \lambda_3 + \lambda_4 + 2\lambda_5 + s_2^- = 2 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 - s^+ = 1 \\ & s_1^- \leq 1 \\ & s_2^- \leq 2\beta_2 \end{aligned}$$

with all variables nonnegative and the condition $s^+ \leq \infty$ omitted. Setting $\beta_2 = 0$ yields $\lambda_1^* = 1$ and all other variables at zero so P_1 is a member of the active solution set and is also characterized as efficient with all slacks at zero in this optimum. Changing to $\beta_2 = 1/8$ yields the following optimum

$$\lambda_1^* = \lambda_2^* = 1/2, s_2^{-*} = 1/4$$

with all other variables at zero. Hence P_1 changes to inefficient status under definition 7.2. This change of status shows that the preceding theorem does not apply in this case because P_1 is needed in the expression of this optimum. In fact, changing from $\beta_2 = 1/8$ to $\beta_2 = 1/4$ gives an optimum with $\lambda_2^* = 1$, $s_2^{-*} = 1/2$ and all other variables zero. P_1 is no longer an active member of this optimum because it is no longer needed with this degree of discretion allowed.

Problem 7.7

Now use model (7.1), the Banker-Morey model, to evaluate P_1 in Figure 7.1. Then interpret and compare the results from this model with the last results secured from model (7.2), the Charnes-Cooper-Rousseau-Semple model, in the preceding question.

Suggested response : Applying (7.1) to the data in Figure 7.1 yields the following model

$$\min \quad \theta - \varepsilon(s_1^- + s^+)$$

$$\begin{aligned}
 \text{subject to} \quad & \theta = \lambda_1 + \lambda_2 + 2\lambda_3 + 3\lambda_4 + 2\lambda_5 + s_1^- \\
 & 2 = 2\lambda_1 + 3/2\lambda_2 + \lambda_3 + \lambda_4 + 2\lambda_5 + s_2^- \\
 & 1 = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 - s^+ \\
 & 0 \leq \lambda_1, \dots, \lambda_5, s_1^-, s_2^-, s^+.
 \end{aligned}$$

An optimal solution to this problem is

$$\theta^* = 1, \lambda_2^* = 1, s_2^{-*} = 1/2$$

and all other variables are zero.

This solution is the same as the one previously secured with model (7.2) but its interpretation is different because no penalty is associated with s_2^{-*} as would be appropriate if this unutilized capacity was left idle because no alternative use for this resource was available. P_1 was characterized as inefficient with this same nonzero slack of $s_2^{-*} = 1/2$ because this value, 25% of this input, could have been used in a less valuable secondary usage. This secondary usage could have been accommodated, moreover, without worsening efficiency in the primary activity undertaken. Hence, in this case it is proper to penalize P_1 for its failure to make this nonzero slack available to this secondary use.

Problem 7.8

The Banker-Morey formulation of non-discretionary variables is described by (7.1) at the beginning of this chapter and the BND model in DEA-Solver can solve the bounded-variable model (7.10) in Section 7.3.

Assignment :

- (1) How can you solve (7.1) by transforming it into (7.10)?
- (2) Solve the example depicted in Figure 7.1 by the BND-I-C (input-oriented bounded variable model under constant returns-to-scale assumption) code.

Suggested Answer :

- (1) Since the non-discretionary inputs in (7.1) are bounded above by the observed values, we can transform this problem into (7.10) by setting

$$l_o^{N_x} = \mathbf{0}, u_o^{N_x} = \mathbf{x}_o^N, L = 0, U = \infty, Y^N = \emptyset \text{ (empty).}$$

- (2) The input data format for the BND model is explained in Appendix B (Section B.5.4). Following this, the input data for the problem described in Figure 7.1 must be prepared on an Excel Sheet as follows,

Problem 7.8	(I)x1	(IB)x2	(LB)x2	(UB)x2	(O)y
P1	1	2	0	2	1
P2	1	1.5	0	1.5	1
P3	2	1	0	1	1
P4	3	1	0	1	1
P5	2	2	0	2	1

Here “(IB)x2” designates x_2 as a bounded input variable, “(LB)x2” and “(UB)x2” denote its lower and upper bounds, respectively.

After running the BND-I-C code, we found the optimal solution as summarized below:

For P_1 : $\theta^* = 1, s_2^- = 1/2$ and is inefficient. Its reference is P_2 .

For P_2 : $\theta^* = 1$ and is efficient.

For P_3 : $\theta^* = 1$ and is efficient.

For P_4 : $\theta^* = 2/3, s_1^- = 1$ and is inefficient. Its reference is P_3 .

For P_5 : $\theta^* = 1/2, s_1^- = 1, s_2^- = 1/2$ and is inefficient. Its reference is P_2 .

Notice that the BND model maximizes the sum of slacks including non-discretionary variables and hence P_5 's projection is P_2 .

Problem 7.9

In Section 7.3.2 we introduced bounded variable models where it is implicitly assumed that the set of controllable variables (C) and that the set of bounded variables (N) are mutually exclusive (disjoint), i.e. $C \cap N = \emptyset$.

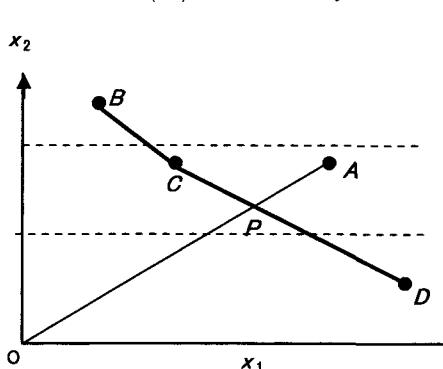


Figure 7.6. Case 1

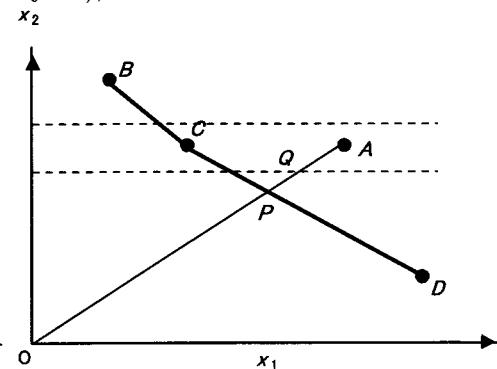


Figure 7.7. Case 2

Consider the case $C \cap N \neq \emptyset$, i.e. some (or all) bounded variables are directly related in the evaluation of the DMU efficiency scores, although they are still bounded.

Assignment : What is induced by this change of situation?

Suggested Response : We illustrate this in the case of two inputs (x_1, x_2) and one “unitized” output ($y = 1$) where it is assumed that $C = \{x_1, x_2\}$ and $N = \{x_2\}$, i.e. x_2 is bounded but should nevertheless be accounted for directly in the efficiency evaluation. We observe two feasible cases as depicted in Figures 7.6 and 7.7. In both cases the efficient frontiers are spanned by bold line segments connecting B, C and D and we evaluate the efficiency of DMU A whose upper and lower bounds of x_2 are designated by broken lines.

Since we incorporated two inputs into evaluating the efficiency of A using an input-oriented model, the CCR projection is the point P on the frontier. In Figure 7.6, the x_2 -value of P is in between the bounds and hence we have $\theta_{CCR}^* = \theta_{BND}^*$ for A . In case 2 in Figure 7.7, on the other hand, the projected point P is outside the bounds and A must therefore be projected to Q , a point which is not on the frontier defined by B , C and D , in order to satisfy the lower bound. Thus, in this case we have $\overline{OP}/\overline{OA} = \theta_{CCR}^* < \theta_{BND}^* = \overline{OQ}/\overline{OA}$ for A . However, these relations will differ by the selection of the sets C and N and hence this inequality need not always be true.

Problem 7.10

In succession to the preceding problem:

- (i) Use the BND-I-C code in DEA-Solver for solving the following problem with two inputs (x_1, x_2) and single unitized output (y), where x_2 is discretionary but bounded and should be included in the evaluation of the efficiency score, i.e. the sets $C = \{x_1, x_2\}$ and $N = \{x_2\}$.

	INPUT		Lower bound	Upper bound	OUTPUT
	x_1	x_2	x_2	x_2	y
A	4	5	4.5	7	1
B	1	8	0	10	1
C	2	6	0	10	1
D	5	2	0	10	1

Thus, A 's efficiency θ^* is evaluated by solving the following LP,

$$\begin{aligned}
 \min \quad & \theta \\
 \text{subject to} \quad & 4\theta = 4\lambda_1 + \lambda_2 + 2\lambda_3 + 5\lambda_4 + s_1^- \quad (\text{for } x_1) \\
 & 5\theta = 5\lambda_1 + 8\lambda_2 + 6\lambda_3 + 2\lambda_4 + s_2^- \quad (\text{for } x_2) \\
 & 1 = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - s^+ \quad (\text{for } y) \\
 & 4.5 \leq 5\lambda_1 + 8\lambda_2 + 6\lambda_3 + 2\lambda_4 \leq 7. \quad (\text{bounds of } x_2 \text{ for } A)
 \end{aligned} \tag{7.24}$$

- (ii) Compare the results of (i) with the case where x_1 is the only controllable variable and x_2 is bounded, i.e. the sets $C = \{x_1\}$ and $N = \{x_2\}$.

In this case (7.24), which connects x_2 directly to the efficiency evaluation, is dropped from the above constraints for evaluating the efficiency of A .

Suggested Response :

- (i) Prepare an Excel sheet as described below.

(1)	(I)x1	(I)x2	(IB)x2	(LB)x2	(UB)x2	(O)y
A	4	5	5	4.5	7	1
B	1	8	8	0	10	1
C	2	6	6	0	10	1
D	5	2	2	0	10	1

Here the variable x_2 appears as an input (I) and as a bounded (IB) variable. This kind of mixed (joint) use is permitted in the BND code.

The solution is summarized as follows,

The DMU A has $\theta^* = 0.9$ and its projection onto the efficient frontier is $x_1 = 3.125$, $x_2 = 4.5$ (the lower bound). All other DMUs are efficient. (This corresponds to Case 2 in Figure 7.7.)

(ii) In this case the Excel sheet is standard, like the one explained in Section B.5.4, and is as follows,

(2)	(I)x1	(IB)x2	(LB)x2	(UB)x2	(O)y
A	4	5	4.5	7	1
B	1	8	0	10	1
C	2	6	0	10	1
D	5	2	0	10	1

The results show that only B is efficient and all other DMUs are inefficient, i.e., $\theta_A^* = 0.375(\lambda_B^* = 1), \theta_C^* = 0.5(\lambda_B^* = 1), \theta_D^* = 0.2(\lambda_B^* = 1)$. Although the projections of C and D are directed to B , A is projected to $x_1 = 1.5, x_2 = 7, y = 1$. This is caused by the upper bound (7) on x_2 listed under (UB)x2 for A in the above table.

Figure 7.8 depicts the relationship between the “primary” input x_1 and output y . From this figure we can see that the efficient frontier is spanned by B

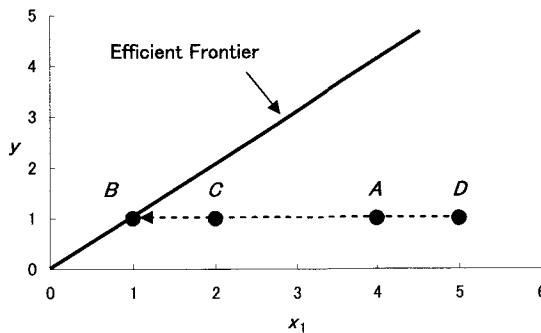


Figure 7.8. Controllable Input and Output

and all other DMUs are to be projected to B with coordinates ($x_1 = 1, x_2 = 8$) in order to become efficient. However, A cannot attain B since its upper bound of $x_2 = 7$ bars it from doing so.

As observed in this example, the choice of “primary” (discretionary) and “secondary” (bounded) variables affects the evaluation of efficiency substantially so care needs to be exercised in deciding which variables should be “primary.”

Problem 7.11 (Categorical Levels Expanded)

In Section 7.4, we introduced the use of “categories” into efficiency evaluations. For the supermarket example discussed in the beginning of Section 7.4, we classified stores facing severe competition as *category* 1 (C_1), in a normal situation as *category* 2 (C_2) and in an advantageous situation as *category* 3 (C_3). Then we evaluated stores in C_1 only with respect to stores within the group. Stores in C_2 were then evaluated with reference to stores in C_1 and C_2 . Finally, stores in C_3 were evaluated with all stores in the entire data set.

Generally, if DMUs in category C_i are used for evaluating DMUs in category C_j , we say C_i is at a *lower level* than C_j and denote this by using the symbol “ \succ ” as in

$$C_i \succ C_j.$$

The above example has level relations,

$$C_1 \succ C_2 \succ C_3.$$

This relation can be diagrammed by Figure 7.9.

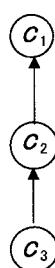


Figure 7.9. Categorization (1)

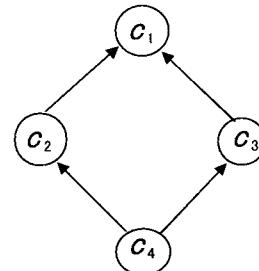


Figure 7.10. Categorization (2)

However, there are situations where the order of levels between categories cannot be readily identified. For example suppose we have four categories of book stores, i.e. stores with both magazine drive-through and free delivery service (C_4), stores with magazine drive-through but no free delivery service (C_2), stores with free delivery service but no magazine drive-through (C_3) and stores with neither magazine drive-through nor free delivery service (C_1). In this case the hierarchy level of C_2 and C_3 is not clearly identified. Hence the level relations can be exhibited by a diagram as Figure 7.10.

Assignment : Enumerate the reference group for each category group in Figure 7.10.

Suggested Response :

Category	Reference Category			
C_1	C_1	C_2	C_3	C_4
C_2		C_2	C_4	
C_3		C_3	C_4	
C_4		C_4		

By this is meant that C_1 can be used to evaluate all 4 reference categories. C_2 can be used to evaluate C_2 and C_4 . C_3 can be used to evaluate C_3 and C_4 . Finally, however, C_4 can be used to evaluate DMUs in only its own class. Thus we have

$$C_1 \succ C_2, C_1 \succ C_3, C_1 \succ C_4, C_2 \succ C_4, C_3 \succ C_4$$

which reduces to $C_1 \succ C_2 \succ C_4$ and $C_1 \succ C_3 \succ C_4$ but no further reduction is possible because no such relation holds for C_2 and C_3 .

Problem 7.12 (Bilateral Comparisons)

Background : In Section 7.3.3 we introduced the data for the twelve Japanese professional baseball teams (see Table 7.3), with the first six teams belonging to “Central League” and the last six to “Pacific League.” It is sometimes said that the Pacific League is not as attractive as the Central League among baseball fans.

Assignment : Use the Bilateral Comparisons model in DEA-Solver to verify this rumor by taking the average salary of “Manager” and “Player” as inputs and “Attendance” as output.

Suggested Answer : Prepare the following Excel sheet as the input data for bilateral comparisons.

Team	(I)Manager	(I)Player	(O)Attendance	League
Swallows	5250	6183	80.87	1
Dragons	2250	5733	93.76	1
Giants	6375	8502	100.00	1
Tigers	3125	4780	76.53	1
Bay Stars	3500	4042	79.12	1
Carp	3125	5623	51.95	1
Lions	5500	10180	56.37	2
Fighters	3625	5362	57.44	2
Blue Wave	2715	4405	58.78	2
Buffalos	3175	6193	53.00	2
Marines	2263	5013	43.47	2
Hawks	3875	3945	82.78	2

where column heading (I) refers to input and (O) refers to output while “League” 1 corresponds to Central and 2 to Pacific, as listed on the right. The “Bilateral” code in DEA-Solver gives the following results,

Rank	DMU	Score	League
1	Dragons	1.924754452	1
2	Tigers	1.133467367	1
3	Hawks	1.07198435	2
4	Bay Stars	1.05497859	1
5	Carp	0.767848928	1
6	Blue Wave	0.746425041	2
7	Giants	0.729163394	1
8	Swallows	0.71851069	1
9	Fighters	0.585915665	2
10	Marines	0.517530223	2
11	Buffalos	0.498371459	2
12	Lions	0.31897526	2

Thus, the rank sum of Central League is 27, while that of Pacific League is 51. The test statistic T of (7.14) is given by

$$T = \frac{27 - 6(6 + 6 + 1)/2}{\sqrt{6 \times 6(6 + 6 + 1)/12}} = -1.921537845.$$

The corresponding significance level is $\alpha = 0.05466$ so we reject the null hypothesis that the two leagues have the same distribution of efficiency scores — at a significance level of 5.466%.

On this basis we say that the Pacific League is less efficient than the Central League in terms of salaries (of managers and players) vs. attendance. This seems consistent with the assertion that the Pacific League is less attractive because this league uses more resources to service its attendance.

Notes

1. R.D. Banker and R. Morey (1986a), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* 34(4), pp.513-521
2. A. Charnes, W.W. Cooper, J.J. Rousseau and J. Semple (1987), "Data Envelopment Analysis and Axiomatic Notions of Efficiency and Reference Sets," CCS Research Report 558 (Austin, Texas: University of Texas, Graduate School of Business, Center for Cybernetic Studies, Austin, Texas 78712).
3. H. Deng (2003), "Evaluating and Managing Congestion in Chinese Production," Ph.D. Thesis (Austin, TX: The Red McCombs School of Business in The University of Texas at Austin). Also available from University Micro Films, Inc., Ann Arbor, Mich.
4. Exceptions are the FDH and multiplicative models discussed in Chapter 4 and Chapter 5.
5. E.F. Fama and M.C. Jensen (1983a), "Separation of Ownership and Control," *Journal of Law and Economics* 26, pp.301-325. See also Fama and Jensen (1983b), "Agency Problems and Residual Claims," *Journal of Law and Economics* 26, pp.327-349.
6. P.L. Brockett, W.W. Cooper, J.J. Rousseau and Y. Wang (1998), "DEA Evaluations of the Efficiency of Organizational Forms and Distribution Systems in the U.S. Property and Liability Insurance Industry," *International Journal of System Science* 29, pp.1235-1247.

7. Solvency, defined as an ability to meet claims, for example, played a critically important role in Brockett *et al.* but not in the Fama-Jensen study. See P.L. Brockett, W.W. Cooper, L. Golden, J.J. Rousseau and Y. Wang "Evaluating Solvency and Efficiency Performances by U.S. Property-Liability Insurance Companies," *Journal of Risk and Insurance* (submitted).
8. The Mann-Whitney statistic does not require equal sample size. See E.L. Lehman (1979) *Nonparametrization: Statistical Methods Based on Ranks* (New York: Holden-Day).
9. D.L. Adolphson, G.C. Cornia, L.C. Walters, "A Unified Framework for Classifying DEA Models," *Operational Research '90*, edited by H.E. Bradley, Pergamon Press, pp.647-657 (1991).
10. K. Tone (1993), "On DEA Models," (in Japanese) *Communications of the Operations Research Society of Japan* 38, pp.34-40.
11. R.D. Banker and R.C. Morey (1986b), "The Use of Categorical Variables in Data Envelopment Analysis," *Management Science* 32(12), pp.1613-1627.
12. W.A. Kamakura (1988), "A Note on The Use of Categorical Variables in Data Envelopment Analysis," *Management Science* 34(10), pp.1273-1276.
13. J.J. Rousseau and J. Semple (1993), "Categorical Outputs in Data Envelopment Analysis," *Management Science* 39(3), pp.384-386.
14. K. Tone (1997), "DEA with Controllable Category Levels," in *Proceedings of the 1997 Spring National Conference of the Operations Research Society of Japan* pp.126-127.
15. K. Tone (1995), "DEA Models Revisited," *Communications of the Operations Research Society of Japan* 40, pp.681-685 (in Japanese).
16. K. Tone (1993), *Data Envelopment Analysis* (in Japanese) (Tokyo:JUSE Press, Ltd.).
17. A. Charnes, W.W. Cooper and E. Rhodes (1981), "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Management Science* 27, pp.668-697.
18. P.L. Brockett and B. Golany (1996), "Using Rank Statistics for Determining Programmatic Efficiency Differences in Data Envelopment Analysis," *Management Science* 42, pp.466-472.
19. P.L. Brockett and A. Levine (1984) *Statistics and Probability and Their Applications* (Fort Worth, TX. CBS College Publishing).

8 ALLOCATION MODELS

8.1 INTRODUCTION

The preceding chapters focused on the technical-physical aspects of production for use in situations where unit price and unit cost information are not available, or where their uses are limited because of variability in the prices and costs that might need to be considered. This chapter turns to the topic of “allocative efficiency” in order to show how DEA can be used to identify types of inefficiency which can emerge for treatment when information on prices and costs are known exactly. Technology and cost are the wheels that drive modern enterprises; some enterprises have advantages in terms of technology and others in cost. Hence, the management is eager to know how and to what extent their resources are being effectively and efficiently utilized, compared to other similar enterprises in the same or a similar field.

Regarding this subject, there are two different situations: one with common unit prices and costs for all DMUs and the other with different prices and costs from DMU to DMU. Section 2 of this chapter deals with the former case. However, the common price and cost assumption is not always valid in actual business and it is demonstrated that efficiency measures based on this assumption can be misleading. So we introduce a new cost-efficiency related model along with new revenue and profit efficiency models in Section 3. Section 4 develops a new formula for decomposition of the observed actual cost based on the new cost efficiency model. Using this formula, we can decompose actual

cost as the sum of the minimum cost and the losses due to technical, price and allocative inefficiencies.

8.2 OVERALL EFFICIENCY WITH COMMON PRICES AND COSTS

8.2.1 Cost Efficiency

Figure 8.1 introduces concepts dealing with “allocative efficiency” that can be traced back to M.J. Farrell (1957) and G. Debreu (1951)¹ who originated many of the ideas underlying DEA. Färe, Grosskopf and Lovell (1985)² developed linear programming formulations of these concepts.

The solid lines in this figure are segments of an isoquant that represents all possible combinations of the input amounts (x_1, x_2) that are needed to produce the same amount of a single output. P is a point in the interior of the production possibility set representing the activity of a DMU which produces this same amount of output but with greater amounts of both inputs.³

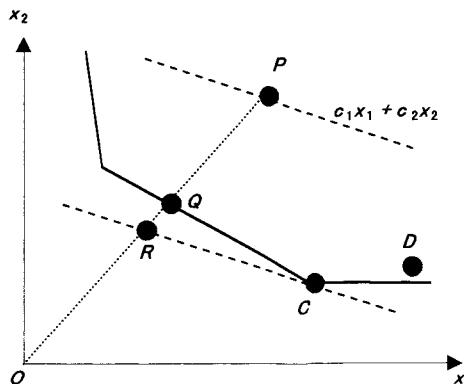


Figure 8.1. Technical, Allocative and Overall Efficiency

To evaluate the performance of P we can use the customary Farrell measure of radial efficiency. Reverting to the notation of Section 1.4, Chapter 1, we can represent this measure in ratio form as

$$0 \leq \frac{d(O, Q)}{d(O, P)} \leq 1,$$

and interpret this as the distance from O to Q relative to the distance from O to P . The result is the measure of technical efficiency that we have customarily represented as θ_o^* .

The components of this ratio lie on the dotted line from the origin through Q to P . To bring price-cost and, hence, “allocative efficiency” considerations into

the picture we turn to the broken line passing through P for which the budget (or cost) line is associated with $c_1x_1 + c_2x_2 = k_1$. However, this cost can be reduced by moving this line in parallel fashion until it intersects the isoquant at C . The coordinates of C then give $c_1x_1^* + c_2x_2^* = k_0$ where $k_0 < k_1$ shows the amount by which total cost can be reduced. Further parallel movement in a downward direction is associated with reduced output so the position of the broken line passing through C is minimal at the prescribed output level. This optimal point C is obtained as the optimal solution \mathbf{x}^* of the following LP (Farrell (1957)):

$$\begin{aligned} [\text{Cost}] \quad & \mathbf{c}\mathbf{x}^* = \min_{\mathbf{x}, \boldsymbol{\lambda}} \mathbf{c}\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \geq X\boldsymbol{\lambda} \\ & \mathbf{y}_o \leq Y\boldsymbol{\lambda} \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \end{aligned} \tag{8.1}$$

where $\mathbf{c} = (c_1, \dots, c_m)$ is the common unit input-price or unit-cost vector.

Now we note that we can similarly determine the relative distances of R and Q to obtain the following ratio,

$$0 \leq \frac{d(O, R)}{d(O, Q)} \leq 1.$$

Farrell refers to this as a measure of “price efficiency” but the more commonly used term is “allocative efficiency.” In either case it provides a measure of the extent to which the technically efficient point, Q , falls short of achieving minimal cost because of failure to make the substitutions (or reallocations) involved in moving from Q to C along the efficiency frontier.

There is one further measure that is commonly referred to as “overall efficiency,” or “cost efficiency.” We can represent this by means of the following ratio,

$$0 \leq \frac{d(O, R)}{d(O, P)} = \frac{\mathbf{c}\mathbf{x}^*}{\mathbf{c}\mathbf{x}_o} \leq 1. \tag{8.2}$$

This is a measure of the extent to which the originally observed values at P , represented in the denominator, have fallen short of achieving the minimum cost represented in the numerator.

To put this in a way that relates all three of these efficiency concepts to each other, note that

$$\frac{d(O, R)}{d(O, Q)} \cdot \frac{d(O, Q)}{d(O, P)} = \frac{d(O, R)}{d(O, P)}. \tag{8.3}$$

In sum, “overall (cost) efficiency” (on the right) is equal to the product of “allocative” times “technical efficiency,” (on the left).

Furthermore, the technical efficiency can be decomposed into the pure technical efficiency and the scale efficiency as defined in Section 4.5 of Chapter 4. Thus, we have the following decomposition:

$$\text{Overall Eff.} = \text{Allocative Eff.} \times \text{Pure Technical Eff.} \times \text{Scale Eff..} \tag{8.4}$$

Or,

$$\text{OE} = \text{AE} \times \text{TE} = \text{AE} \times \text{PTE} \times \text{SE}. \quad (8.5)$$

8.2.2 Revenue Efficiency

Given the common unit price vector $\mathbf{p} = (p_1, \dots, p_s)$ for the output \mathbf{y} , we evaluate the revenue efficiency of DMU_o as follows:

$$\begin{aligned} [\text{Revenue}] \quad & \mathbf{p}\mathbf{y}^* = \max_{\mathbf{y}, \boldsymbol{\lambda}} \mathbf{p}\mathbf{y} \\ \text{subject to} \quad & \mathbf{x}_o \geq X\boldsymbol{\lambda} \\ & \mathbf{y} \leq Y\boldsymbol{\lambda} \\ & L \leq \mathbf{e}\boldsymbol{\lambda} \leq U \\ & \boldsymbol{\lambda} \geq 0. \end{aligned} \quad (8.6)$$

We inserted an additional constraint on scale ($L \leq \mathbf{e}\boldsymbol{\lambda} \leq U$) in order to cope with various returns-to-scale assumptions.

This model allows substitutions in outputs. Let the optimal solution be $(\mathbf{y}^*, \boldsymbol{\lambda}^*)$. Then, the *revenue efficiency* is defined in ratio form as:

$$E_R \text{ (Revenue Efficiency)} = \frac{\mathbf{p}\mathbf{y}_o}{\mathbf{p}\mathbf{y}^*}. \quad (8.7)$$

We have $0 \leq E_R \leq 1$ and $\text{DMU}(\mathbf{x}_o, \mathbf{y}_o)$ is *revenue efficient* if and only if $E_R = 1$.

8.2.3 Profit Efficiency

To express the profit of DMU_o we use the common unit price vector \mathbf{p} and unit cost vector \mathbf{c} , to obtain the following LP problem:

$$\begin{aligned} [\text{Profit}] \quad & \mathbf{p}\mathbf{y}^* - \mathbf{c}\mathbf{x}^* = \max_{\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}} \mathbf{p}\mathbf{y} - \mathbf{c}\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} = X\boldsymbol{\lambda} \leq \mathbf{x}_o \\ & \mathbf{y} = Y\boldsymbol{\lambda} \geq \mathbf{y}_o \\ & L \leq \mathbf{e}\boldsymbol{\lambda} \leq U \\ & \boldsymbol{\lambda} \geq 0. \end{aligned} \quad (8.8)$$

This formulation extends (8.1) with an additional constraint on scale ($L \leq \mathbf{e}\boldsymbol{\lambda} \leq U$). So, substitutions in inputs or outputs are not allowed in this case. Here, the purpose is to find a profit-maximization mix in the production possibility set $P = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, L \leq \mathbf{e}\boldsymbol{\lambda} \leq U, \boldsymbol{\lambda} \geq 0\}$. Based on an optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$, the *profit efficiency* can be defined in ratio form by

$$E_P \text{ (Profit Efficiency)} = \frac{\mathbf{p}\mathbf{y}_o - \mathbf{c}\mathbf{x}_o}{\mathbf{p}\mathbf{y}^* - \mathbf{c}\mathbf{x}^*}. \quad (8.9)$$

where \mathbf{y}^* , \mathbf{x}^* are optimal for (8.8) and $\mathbf{y}_o, \mathbf{x}_o$ are the vectors of observed values for DMU_o.

Under the assumption $p\mathbf{y}_o > c\mathbf{x}_o$, we have $0 < E_P \leq 1$ and DMU (\mathbf{x}_o , \mathbf{y}_o) is *profit efficient* if and only if $E_P = 1$. The differences between \mathbf{x}^* and \mathbf{x}_o and between \mathbf{y}^* and \mathbf{y}_o may suggest directions for managerial improvement and this can be analyzed, constraint by constraint, in (8.8).

8.2.4 An Example

The data in the following table will provide examples to illustrate the use of these models. Here each of three DMUs produces a single output in the amount y , shown in the column for y under "Output," by using two inputs in the amounts x_1 and x_2 shown in the two columns headed by x_1 and x_2 . The common unit costs and price are exhibited in the columns for c_1 , c_2 and p , respectively.

Table 8.1. Sample Data for Allocative Efficiency

DMU	Input				Output	
	x_1	c_1	x_2	c_2	y	p
A	3	4	2	2	3	6
B	1	4	3	2	5	6
C	4	4	6	2	6	6

We solved this data set under the constant returns-to-scale assumption so the constraint $L \leq e\lambda \leq U$ was omitted. We then obtained the results shown in Table 8.2. DMU B is the only one that is efficient and is the best performer in all efficiency measures.

Table 8.2. Efficiencies

DMU	Efficiency				
	Technical	Cost	Allocative	Revenue	Profit
A	0.9	0.375	0.417	0.9	0.15
B	1	1	1	1	1
C	0.6	0.429	0.715	0.6	0.2

8.3 NEW COST EFFICIENCY UNDER DIFFERENT UNIT PRICES

Firstly we observe an unacceptable property of the traditional Farrell-Debreu cost efficiency models described in the preceding section which can occur when the unit prices of input are not identical among DMUs.

Suppose that DMUs A and B have the same amount of inputs and outputs, i.e., $\mathbf{x}_A = \mathbf{x}_B$ and $\mathbf{y}_A = \mathbf{y}_B$. Assume further that the unit cost of DMU A is twice that of DMU B for each input, i.e., $\mathbf{c}_A = 2\mathbf{c}_B$. Under these assumptions, we have the following theorem:

Theorem 8.1 (Tone(2002)⁴) *Using the Farrell-Debreu cost efficiency model both DMUs A and B have the same cost (overall) and allocative efficiencies even when the latter is more costly than the former.*

Proof: Since DMUs A and B have the same inputs and outputs, they have the same technical efficiency, i.e., $\theta_A^* = \theta_B^*$. The Farrell measure of cost efficiency for DMU A (or DMU B) can be obtained by solving the following LP (see (8.1)):

$$\min \mathbf{c}_A \mathbf{x} (= 2\mathbf{c}_B \mathbf{x}) \quad (8.10)$$

$$\text{subject to } x_i \geq \sum_{j=1}^n x_{ij} \lambda_j \quad (i = 1, \dots, m) \quad (8.11)$$

$$y_{rA} (= y_{rB}) \leq \sum_{j=1}^n y_{rj} \lambda_j \quad (r = 1, \dots, s) \quad (8.12)$$

$$\lambda_j \geq 0. \quad (\forall j) \quad (8.13)$$

Apparently, DMUs A and B have the same optimal solution (inputs) $\mathbf{x}_A^* = \mathbf{x}_B^*$, and hence the same cost efficiency, since we have:

$$\gamma_A^* = \mathbf{c}_A \mathbf{x}_A^* / \mathbf{c}_A \mathbf{x}_A = 2\mathbf{c}_B \mathbf{x}_B^* / 2\mathbf{c}_B \mathbf{x}_B = \mathbf{c}_B \mathbf{x}_B^* / \mathbf{c}_B \mathbf{x}_B = \gamma_B^*.$$

By definition, they also have the same allocative efficiency. \square

This is not acceptable, since DMUs A and B have the same cost and allocative efficiencies but the cost of DMU B is half that of DMU A.

8.3.1 A New Scheme for Evaluating Cost Efficiency

The previous example reveals a serious shortcoming in the traditional Farrell-Debreu cost and allocative efficiency measures. These shortcomings are caused by the structure of the supposed production possibility set P as defined by:

$$P = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}. \quad (8.14)$$

P is defined only by using technical factors $X = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in R^{m \times n}$ and $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n) \in R^{s \times n}$, but excludes consideration of the unit input costs $C = (\mathbf{c}_1, \dots, \mathbf{c}_n)$.

Let us define another cost-based production possibility set P_c as:

$$P_c = \{(\bar{x}, y) | \bar{x} \geq \bar{X}\lambda, y \leq Y\lambda, \lambda \geq 0\}, \quad (8.15)$$

where $\bar{X} = (\bar{x}_1, \dots, \bar{x}_n)$ with $\bar{x}_j = (c_{1j}x_{1j}, \dots, c_{mj}x_{mj})^T$.

Here we assume that the matrices X and C are non-negative. We also assume that the elements of $\bar{x}_{ij} = (c_{ij}x_{ij})$ ($\forall(i, j)$) are denominated in homogeneous units, viz., dollars, so that adding up the elements of \bar{x}_{ij} has a well defined meaning.

Based on this new production possibility set P_c , a new “technical efficiency” measure, $\bar{\theta}^*$, is obtained as the optimal solution of the following LP problem:

$$[\text{NTech}] \quad \bar{\theta}^* = \min_{\bar{\theta}, \lambda} \bar{\theta} \quad (8.16)$$

$$\text{subject to} \quad \bar{\theta}\bar{x}_o \geq \bar{X}\lambda \quad (8.17)$$

$$y_o \leq Y\lambda \quad (8.18)$$

$$\lambda \geq 0. \quad (8.19)$$

The new cost efficiency $\bar{\gamma}^*$ is defined as

$$\bar{\gamma}^* = e\bar{x}_o^*/e\bar{x}_o, \quad (8.20)$$

where $e \in R^m$ is a row vector with all elements being equal to 1, and \bar{x}_o^* is the optimal solution of the LP given below:

$$[\text{NCost}] \quad e\bar{x}_o^* = \min_{\bar{x}, \lambda} e\bar{x} \quad (8.21)$$

$$\text{subject to} \quad \bar{x} \geq \bar{X}\lambda \quad (8.22)$$

$$y_o \leq Y\lambda \quad (8.23)$$

$$\lambda \geq 0. \quad (8.24)$$

Theorem 8.2 *The new “cost efficiency”, $\bar{\gamma}^*$ in (8.20), is not greater than the new “technical efficiency” $\bar{\theta}^*$ in (8.16).*

Proof : Let an optimal solution for (8.16)-(8.19) be $(\bar{\theta}^*, \lambda^*)$. Then, $(\bar{\theta}^*\bar{x}_o, \lambda^*)$ is feasible for (8.21)-(8.24). Hence, it follows that $e\bar{\theta}^*\bar{x}_o \geq e\bar{x}_o^*$. This leads to $\bar{\theta}^* \geq e\bar{x}_o^*/e\bar{x}_o = \bar{\gamma}^*$. \square

The new *allocative* efficiency $\bar{\alpha}^*$ is then defined as the ratio of $\bar{\gamma}^*$ to $\bar{\theta}^*$, i.e.,

$$\bar{\alpha}^* = \bar{\gamma}^*/\bar{\theta}^*. \quad (8.25)$$

We note that the new efficiency measures $\bar{\theta}^*$, $\bar{\gamma}^*$ and $\bar{\alpha}^*$ are all units invariant so long as \bar{X} has a common unit of cost, e.g., dollars, cents or pounds.

On monotonicity of the new measures with respect to unit cost, we have the following theorem.

Theorem 8.3 (Tone(2002)) If $x_A = x_B$, $y_A = y_B$ and $c_A \geq c_B$, then we have the following inequalities: $\bar{\theta}_A^* \leq \bar{\theta}_B^*$ and $\bar{\gamma}_A^* \leq \bar{\gamma}_B^*$. Furthermore, strict inequalities hold if $c_A > c_B$.

Proof. Since $\bar{x}_A \geq \bar{x}_B$ and $y_A = y_B$, the new technical measure $\bar{\theta}_A^*$ is less than or equal to $\bar{\theta}_B^*$ and a strict inequality holds if $c_A > c_B$. Regarding the new cost efficiency, we note that the optimal solution of [NCost] depends only on y_o . Hence, DMUs (\bar{x}_A, y_A) and (\bar{x}_B, y_B) with $y_A = y_B$ have a common optimal solution \bar{x}^* . Therefore, we have $\bar{\gamma}_A^* = e\bar{x}^*/e\bar{x}_A \leq e\bar{x}^*/e\bar{x}_B = \bar{\gamma}_B^*$ and strict inequality holds if $c_A > c_B$. \square

Thus, the new measure eliminates the possible occurrence of the phenomenon observed at the beginning of this section.

8.3.2 Differences Between the Two Models

We now comment on the differences existing between the traditional “Farrell-Debreu” and the new models. In the traditional model, keeping the unit cost of DMU_o fixed at c_o , the optimal input mix x^* that produces the output y_o is found. In the new model, we search for the optimal input mix \bar{x}^* for producing y_o (or more). More concretely, the optimal mix is described as:

$$\bar{x}_i^* = \sum_{j=1}^n c_{ij} x_{ij} \lambda_j^*. \quad (i = 1, \dots, m) \quad (8.26)$$

Hence, it is assumed that, for a given output y_o , the optimal input mix can be found (and realized) independently of the current unit cost c_o of DMU_o.

These are fundamental differences between the two models. Using the traditional “Farrell-Debreu” model we can fail to recognize the existence of other cheaper input mixes, as we have demonstrated earlier. We demonstrate this with a simple example involving three DMUs A, B and C with each using two inputs (x_1, x_2) to produce one output (y) along with input costs (c_1, c_2). The data and the resulting measures are exhibited in Table 8.3.

For DMUs A and B, the traditional model gives the same technical (θ^*), cost (γ^*) and allocative (α^*) efficiency scores — as expected from Theorem 1. DMU C is found to be the only efficient performer in this framework.

The new scheme devised as in Tone (2002) — see footnote 4 — distinguishes DMU A from DMU B by according them different technical and cost efficiency scores. (See New Scheme in Table 8.3). This is due to the difference in their unit costs. Moreover, DMU B is judged as technically, cost and allocatively efficient with improvement in cost efficiency score from $0.35(\gamma_B^*)$ to $1(\bar{\gamma}_B^*)$ as exhibited in these two tables. As shown in Table 8.3, this cost difference produces a drop in DMU A’s cost efficiency score from $0.35(\gamma_A^*)$ to $0.1(\bar{\gamma}_A^*)$. This drop in DMU A’s performance is explained by its higher cost structure. Lastly, DMU C is no longer efficient in any of its technical, cost or allocative efficiency performances.

Table 8.3. Comparison of Traditional and New Scheme

	x_1	c_1	x_2	c_2	y	Traditional Efficiency		
						Tech.	Cost	Alloc.
						θ^*	γ^*	α^*
A	10	10	10	10	1	0.5	0.35	0.7
B	10	1	10	1	1	0.5	0.35	0.7
C	5	3	2	6	1	1	1	1

	\bar{x}_1	e_1	\bar{x}_2	e_2	y	New Scheme Efficiency		
						Tech.	Cost	Alloc.
						$\bar{\theta}^*$	$\bar{\gamma}^*$	$\bar{\alpha}^*$
A	100	1	100	1	1	0.1	0.1	1
B	10	1	10	1	1	1	1	1
C	15	1	12	1	1	0.8333	0.7407	0.8889

8.3.3 An Empirical Example

In this section, we apply our new method to a set of hospital data. Table 8.4 records the performances of 12 hospitals in terms of two inputs, number of doctors and nurses, and two outputs identified as number of outpatients and inpatients (each in units of 100 persons/month). Relative unit costs of doctors and nurses for each hospital are also recorded in columns 4 and 6.

Multiplying the number of doctors and nurses by their respective unit costs we obtain the new data set (\bar{X}, Y) exhibited in Table 8.5. The results of efficiency scores: CCR(θ^*), New technical ($\bar{\theta}^*$), New cost ($\bar{\gamma}^*$) and New allocative ($\bar{\alpha}^*$), are also recorded.

From the results, it is seen that the best performer is Hospital B with all its efficiency scores being equal to one. Regarding the cost-based measures, Hospitals E and L received full efficiency marks even though they fell short in their CCR efficiency score. Conversely, although E has the worst CCR score (0.763), its lower unit costs are sufficient to move its cost-based performance to the top rank. This information obtained from $\theta^* = 0.763$ shows that this hospital still has room for input reductions compared with other technically efficient hospitals. Hospital L, on the other hand, may be regarded as positioned in the best performer group. These two DMUs show that the usual assumption does not hold and thus technical efficiency (CCR $\theta^* = 1$) being achieved is not a necessary condition for the new cost and allocative efficiencies. This is caused by the difference between the two production possibility sets, i.e., the technology-based (X, Y) and the cost-based (\bar{X}, Y) . On the other hand, Hospital D is rated worst with respect to cost-based measures, although it receives full efficiency marks in terms of its CCR score. This gap is due to its

Table 8.4. Data for 12 Hospitals

No.	DMU	Inputs				Outputs	
		Doctor Number	Cost	Nurse Number	Cost	Outpat. Number	Inpat. Number
1	A	20	500	151	100	100	90
2	B	19	350	131	80	150	50
3	C	25	450	160	90	160	55
4	D	27	600	168	120	180	72
5	E	22	300	158	70	94	66
6	F	55	450	255	80	230	90
7	G	33	500	235	100	220	88
8	H	31	450	206	85	152	80
9	I	30	380	244	76	190	100
10	J	50	410	268	75	250	100
11	K	53	440	306	80	260	147
12	L	38	400	284	70	250	120
Average		33.6	435.8	213.8	85.5	186.3	88.2

Table 8.5. New Data Set and Efficiencies

No.	DM	Data				Efficiency			
		\bar{X}		Y		CCR θ^*	Tech. $\bar{\theta}^*$	Cost $\bar{\gamma}^*$	Alloc. $\bar{\alpha}^*$
		Doctor	Nurse	Inp.	Outp.				
1	A	10000	15100	100	90	1	.994	.959	.965
2	B	6650	10480	150	50	1	1	1	1
3	C	11250	14400	160	55	.883	.784	.724	.923
4	D	16200	20160	180	72	1	.663	.624	.941
5	E	6600	11060	94	66	.763	1	1	1
6	F	24750	20400	230	90	.835	.831	.634	.764
7	G	16500	23500	220	88	.902	.695	.693	.997
8	H	13950	17510	152	80	.796	.757	.726	.959
9	I	11400	18544	190	100	.960	.968	.953	.984
10	J	20500	20100	250	100	.871	.924	.776	.841
11	K	23320	24480	260	147	.955	.995	.863	.867
12	L	15200	19880	250	120	.958	1	1	1

high cost structure. Hospital D needs reductions in its unit costs to attain good cost-based scores.

Hospital F has the worst allocative efficiency score, and hence needs a change in input-(cost)mix. This hospital has the current input-(cost)mix, $\bar{x}_F = (24750, 20400)$, while the optimal mix \bar{x}_F^* is $(11697, 16947)$. So, under its current costs, F needs to reduce the number of doctors from 55 to 26 ($=11697/450$), and nurses from 255 to 212 ($=16947/80$). Otherwise, if F retains its current input numbers, it needs to reduce the unit cost of doctors from 450 to 213 ($=11697/55$), and that of nurses from 80 to 66 ($=16947/255$). Of course, there are many other adjustment plans as well. In any case our proposed new measures provide much more information than the traditional ones.

8.3.4 Extensions

We can also extend this new cost efficiency model to three other situations as follows.

1. Revenue Efficiency

Given the unit price p_j for each output y_j ($j = 1, \dots, n$), the conventional revenue efficiency ρ_o^* of DMU_o is evaluated by $\rho_o^* = p_o y_o / p_o y_o^*$. Here, $p_o y_o^*$ is obtained from the optimal objective value of LP problem (8.6).

However, in the situation of different unit prices, this revenue efficiency ρ_o^* suffers from shortcomings similar to the traditional cost efficiency measure shortcomings described in the previous section. We can eliminate such shortcomings by introducing the price-based output $\bar{Y} = (\bar{y}_1, \dots, \bar{y}_n)$ with $\bar{y}_j = (p_{1j} y_{1j}, \dots, p_{sj} y_{sj})$ into the following LP:

$$\begin{aligned} [\text{NRevenue}] \quad & e\bar{y}_o^* = \max_{\bar{y}, \lambda} e\bar{y} \quad (8.27) \\ \text{subject to} \quad & x_o \geq X\lambda \\ & \bar{y} \leq \bar{Y}\lambda \\ & L \leq e\lambda \leq U \\ & \lambda \geq 0. \end{aligned}$$

The new revenue efficiency measure $\bar{\rho}_o$ is defined by

$$\text{New Revenue Efficiency } (\bar{\rho}_o) = e\bar{y}_o / e\bar{y}_o^*. \quad (8.28)$$

2. Profit Efficiency

Using the new cost and revenue efficiency models, we can also define a new profit efficiency model as follows:

$$[\text{NProfit}] \quad e\bar{y}_o^* - e\bar{x}_o^* = \max_{\bar{x}, \bar{y}, \lambda} e\bar{y} - e\bar{x} \quad (8.29)$$

$$\begin{aligned} \text{subject to} \quad & \bar{x} = \bar{X}\lambda \leq \bar{x}_o \\ & \bar{y} = \bar{Y}\lambda \geq \bar{y}_o \\ & L \leq e\lambda \leq U \\ & \lambda \geq 0, \end{aligned}$$

where \bar{X} and \bar{Y} are defined in the new cost and revenue models, respectively. The new profit efficiency is defined as:

$$\text{New Profit Efficiency } (\bar{\pi}_o) = (\mathbf{e}\bar{y}_o - \mathbf{e}\bar{x}) / (\mathbf{e}\bar{y}_o^* - \mathbf{e}\bar{x}_o^*). \quad (8.30)$$

3. Profit Ratio Model

We also propose a model for maximizing the revenue vs. cost ratio,

$$\frac{\text{revenue}}{\text{expenses}},$$

instead of maximizing profit (revenue – expenses), since in some situations the latter gives a negative value that is awkward to deal with. This new profit ratio model can be formulated as a problem of maximizing the *revenue/expenses* ratio to obtain the following fractional programming problem,⁵

$$\begin{aligned} [\text{Profit Ratio}] \quad & \max_{\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}} \frac{\mathbf{p}_o \mathbf{y}}{\mathbf{c}_o \mathbf{x}} \\ \text{subject to} \quad & \mathbf{x} = \mathbf{X}\boldsymbol{\lambda} \leq \mathbf{x}_o \\ & \mathbf{y} = \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{y}_o \\ & L \leq \mathbf{e}\boldsymbol{\lambda} \leq U \\ & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \quad (8.31)$$

We can transform this program to the linear programming problem below, by introducing a variable $t \in R$ and use the Charnes-Cooper transformation of fractional programming which sets $\hat{\mathbf{x}} = t\mathbf{x}$, $\hat{\mathbf{y}} = t\mathbf{y}$, $\hat{\boldsymbol{\lambda}} = t\boldsymbol{\lambda}$. Then multiplying all terms by $t > 0$, we change (8.31) to

$$\begin{aligned} & \max_{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\boldsymbol{\lambda}}, t} \frac{\mathbf{p}_o \hat{\mathbf{y}}}{\mathbf{c}_o \hat{\mathbf{x}}} \\ \text{subject to} \quad & \mathbf{c}_o \hat{\mathbf{x}} = 1 \\ & t\mathbf{x}_o \geq \mathbf{X}\hat{\boldsymbol{\lambda}} = \hat{\mathbf{x}} \\ & t\mathbf{y}_o \leq \mathbf{Y}\hat{\boldsymbol{\lambda}} = \hat{\mathbf{y}} \\ & Lt \leq \mathbf{e}\hat{\boldsymbol{\lambda}} \leq Ut \\ & \hat{\boldsymbol{\lambda}} \geq \mathbf{0}. \end{aligned} \quad (8.32)$$

Let an optimal solution of this LP problem be $(t^*, \hat{\mathbf{x}}^*, \hat{\mathbf{y}}^*, \hat{\boldsymbol{\lambda}}^*)$. Since $t^* > 0$ we can reverse this transformation and obtain an optimal solution to the fractional program in (8.31) from

$$\mathbf{x}^* = \hat{\mathbf{x}}^*/t^*, \quad \mathbf{y}^* = \hat{\mathbf{y}}^*/t^*, \quad \boldsymbol{\lambda}^* = \hat{\boldsymbol{\lambda}}^*/t^*. \quad (8.33)$$

The *revenue/cost* efficiency E_{RC} of DMU_o can then be related to actual revenue and costs by the following “ratio of ratios,”

$$E_{RC} = \frac{\mathbf{p}_o \mathbf{y}_o / \mathbf{c}_o \mathbf{x}_o}{\mathbf{p}_o \mathbf{y}^* / \mathbf{c}_o \mathbf{x}^*}. \quad (8.34)$$

As noted in the following remark, this efficiency index is related to profit efficiency but is applicable even when there are many deficit-DMUs, i.e. DMU_o s for which $\mathbf{p}_o \mathbf{y}_o - \mathbf{c}_o \mathbf{x}_o < 0$.

[Remark] We can follow Cooper *et al.* (2005)⁶ and obtain a profit-to-cost return ratio by noting that an optimal solution to (8.31) is not altered by replacing the objective with

$$\max_{\mathbf{x}, \mathbf{y}, \lambda} \frac{\mathbf{p}_o \mathbf{y}}{\mathbf{c}_o \mathbf{x}} - 1.$$

This gives

$$\frac{\mathbf{p}_o \mathbf{y}^*}{\mathbf{c}_o \mathbf{x}^*} - \frac{\mathbf{c}_o \mathbf{x}^*}{\mathbf{c}_o \mathbf{x}^*} = \frac{\mathbf{p}_o \mathbf{y}^* - \mathbf{c}_o \mathbf{x}^*}{\mathbf{c}_o \mathbf{x}^*}$$

which is the commonly used profit-to-cost ratio measure of performance, after it has been adjusted to eliminate inefficiencies. Finally, if the observed profit-to-cost ratio is positive we can use the following

$$0 \leq \frac{\mathbf{p}_o \mathbf{y}_o - \mathbf{c}_o \mathbf{x}_o}{\mathbf{c}_o \mathbf{x}_o} \Big/ \frac{\mathbf{p}_o \mathbf{y}^* - \mathbf{c}_o \mathbf{x}^*}{\mathbf{c}_o \mathbf{x}^*} \leq 1$$

as a measure of efficiency with unity achieved if and only if $\frac{\mathbf{p}_o \mathbf{y}_o - \mathbf{c}_o \mathbf{x}_o}{\mathbf{c}_o \mathbf{x}_o} = \frac{\mathbf{p}_o \mathbf{y}^* - \mathbf{c}_o \mathbf{x}^*}{\mathbf{c}_o \mathbf{x}^*}$.

8.4 DECOMPOSITION OF COST EFFICIENCY

Technology and cost are the wheels that drive modern enterprises. Some enterprises have advantages in terms of technology and others in cost. Hence, a management may want to know how and to what extent their resources are being effectively and efficiently utilized, compared to similar enterprises in the same or a similar field.

In an effort to address this subject, Tone and Tsutsui (2004)⁷ developed a scheme for decomposing actual observed cost into the sum of the minimum cost and the loss due to input inefficiency. Furthermore, the loss due to input inefficiency can be expressed as the sum of the loss due to input technical, price and allocative inefficiencies.

8.4.1 Loss due to Technical Inefficiency

We consider n DMUs, each having m inputs for producing s outputs. We utilize the notations for denoting observed inputs (\mathbf{x}_o), outputs (\mathbf{y}_o) and input unit prices (\mathbf{c}_o). We also assume that unit input prices are not identical among DMUs. The actual (observed) input cost for DMU ($\mathbf{x}_o, \mathbf{y}_o$) can be calculated as follows:

$$C_o = \sum_{i=1}^m c_{io} x_{io}. \quad (o = 1, \dots, n) \tag{8.35}$$

We postulate the production possibility set P defined by

$$P = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}. \quad (8.36)$$

Let the (technically) efficient input for DMU_{*o*} be \mathbf{x}_o^* , which can be obtained by solving the CCR, SBM or Hybrid models depending on the situation. The technically efficient input cost for DMU_{*o*} is calculated as

$$C_o^* = \sum_{i=1}^m c_{io} x_{io}^*. \quad (8.37)$$

Then, the loss in input cost due to technical inefficiency is expressed as follows:

$$L_o^* = C_o - C_o^* (\geq 0). \quad (8.38)$$

8.4.2 Loss due to Input Price Inefficiency

We now construct a cost-based production possibility set analogous to that in the preceding section as follows:

$$\bar{P}_c = \{(\bar{\mathbf{x}}, \mathbf{y}) | \bar{\mathbf{x}} \geq \bar{X}\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}, \quad (8.39)$$

where $\bar{X} = (\bar{x}_1, \dots, \bar{x}_n) \in R^{m \times n}$, $\bar{\mathbf{x}}_j = (\bar{x}_{ij}, \dots, \bar{x}_{mj})$, and $\bar{x}_{ij} = c_{ij}x_{ij}^*$. It should be noted that x_j^* represents the technically efficient input for producing y_j . Hence, we utilize $c_{ij}x_{ij}^*$ instead of $c_{ij}x_{ij}$ in the preceding section in order to eliminate technical inefficiency to the maximum possible extent. Then we solve the CCR model on \bar{P}_c in a manner similar to that of [NTech] in (8.16):

$$[\text{NTech-2}] \quad \rho^* = \min_{\rho, \boldsymbol{\mu}, \mathbf{t}^-, \mathbf{t}^+} \rho \quad (8.40)$$

$$\text{subject to} \quad \rho \bar{\mathbf{x}}_o = \bar{X}\boldsymbol{\mu} + \mathbf{t}^- \quad (8.41)$$

$$\mathbf{y}_o = Y\boldsymbol{\mu} - \mathbf{t}^+ \quad (8.42)$$

$$\boldsymbol{\mu} \geq \mathbf{0}, \mathbf{t}^- \geq \mathbf{0}, \mathbf{t}^+ \geq \mathbf{0}. \quad (8.43)$$

Let $(\rho^*, \boldsymbol{\mu}^*, \mathbf{t}^{-*}, \mathbf{t}^{+*})$ be an optimal solution for [NTech-2]. Then, $\rho^* \bar{\mathbf{x}}_o = (\rho^* c_{1o} x_{1o}^*, \dots, \rho^* c_{mo} x_{mo}^*)$ indicates the radially reduced input vector on the (weakly) efficient frontier of the cost-based production set \bar{P}_c in (8.39). Now we define

$$\mathbf{c}_o^* = \rho^* \mathbf{c}_o = (\rho^* c_{1o}, \dots, \rho^* c_{mo}). \quad (8.44)$$

\mathbf{c}_o^* is the radially reduced input factor price vector for the technically efficient input \mathbf{x}_o^* that can produce \mathbf{y}_o . The [NTech-2] projection is given by

$$[\text{NTech-2 Projection}] \quad \bar{\mathbf{x}}_o^* = \rho^* \bar{\mathbf{x}}_o - \mathbf{t}^{-*}, \quad \mathbf{y}_o^* = \mathbf{y}_o + \mathbf{t}^{+*}. \quad (8.45)$$

We define the strongly efficient cost C_o^{**} , which is the technical and price efficient cost, and the loss L_o^{**} due to the difference of the input price as follows:

$$C_o^{**} = \sum_{i=1}^m \bar{x}_{io}^* = \sum_{i=1}^m (\rho^* \bar{x}_{io} - t_{io}^{-*}) \leq \rho^* \sum_{i=1}^m \bar{x}_{io} = \rho^* C_o^* \leq C_o^* \quad (8.46)$$

$$L_o^{**} = C_o^* - C_o^{**} (\geq 0). \quad (8.47)$$

8.4.3 Loss due to Allocative Inefficiency

Furthermore, we solve the [NCost] model in (8.21) for \bar{P}_c as follows:

$$\begin{aligned} [\text{NCost-2}] \quad C_o^{***} &= \min_{\bar{x}, \mu} e\bar{x} \\ \text{subject to} \quad \bar{x} &\geq \bar{X}\mu \\ y_o &\leq Y\mu \\ \mu &\geq 0. \end{aligned} \quad (8.48)$$

Let (\bar{x}_o^{**}, μ^*) be an optimal solution. Then, the cost-based pair (\bar{x}_o^{**}, y_o) is the minimum production cost in the assumed production possibility set \bar{P}_c . This set can differ substantially from P if the unit prices of the inputs vary from DMU to DMU. The (global) allocative efficiency α^* of DMU_o is defined as follows:

$$\alpha^* = \frac{C_o^{***}}{C_o^{**}} (\leq 1). \quad (8.49)$$

We also define the loss L_o^{***} due to the suboptimal cost mix as

$$L_o^{***} = C_o^* - C_o^{***} (\geq 0). \quad (8.50)$$

8.4.4 Decomposition of the Actual Cost

From (8.38), (8.47) and (8.50), we can derive at the following theorem.

Theorem 8.4 (Tone and Tsutsui (2004))

$$C_o \geq C_o^* \geq C_o^{**} \geq C_o^{***}. \quad (8.51)$$

Furthermore, we can obtain the relationship among the optimal cost and losses, and the actual cost (C_o) can be decomposed into three losses and the minimum cost (C_o^{***}):

$$L_o^* = C_o - C_o^* (\geq 0) \text{ Loss due to Technical Inefficiency} \quad (8.52)$$

$$L_o^{**} = C_o^* - C_o^{**} (\geq 0) \text{ Loss due to Price Inefficiency} \quad (8.53)$$

$$L_o^{***} = C_o^{**} - C_o^{***} (\geq 0) \text{ Loss due to Allocative Inefficiency} \quad (8.54)$$

$$C_o = L_o^* + L_o^{**} + L_o^{***} + C_o^{***}. \quad (8.55)$$

For further developments of this scheme, see Problems 8.1-8.4 at the end of this chapter.

8.4.5 An Example of Decomposition of Actual Cost

We applied the above procedure to the data set exhibited in Table 8.4 and obtained the results listed in Table 8.6. We then utilized the input-oriented CCR model and the projection formulas in Chapter 3 for finding the technical efficient inputs (x_o^*).

Table 8.6. Decomposition of Actual Cost

DMU	Cost			Loss		
	Actual <i>C</i>	Minimum <i>C</i> ***	C^{***}/C	Tech. <i>L</i> *	Price <i>L</i> **	Alloc. <i>L</i> ***
A	25100	18386	0.73	0	5959	754
B	17130	17130	1	0	0	0
C	25650	18404	0.72	3557	2658	1032
D	36360	21507	0.59	0	13470	1383
E	17660	13483	0.76	4177	0	0
F	45150	27323	0.61	12911	1767	3149
G	40000	26287	0.66	4256	8931	526
H	31460	19684	0.63	6407	4230	1139
I	29944	24605	0.82	3254	1617	468
J	40600	29871	0.74	7725	0	3004
K	47800	34476	0.72	5367	5030	2927
L	35080	31457	0.90	2348	0	1275
Total	391934	282615	0.72	50002	43661	15656

As can be seen from the results, Hospital B is again the most efficient DMU in the sense that it has no loss due to technical, price or allocative inefficiencies, while Hospital D has the worst ratio $C^{***}/C (= 0.59)$ caused by losses due to price and allocative inefficiencies. Hospital A has inefficiency in cost-based aspects, i.e., price and allocation but not in technical-physical aspects, whereas E has inefficiency due to technical-physical aspects but not in cost-based aspects. Figure 8.2 exhibits the decomposition graphically.

8.5 SUMMARY OF CHAPTER 8

This chapter has covered approaches which have been studied in DEA for evaluations of efficiencies such as “allocative” and “overall” efficiencies. Problems in the standard approaches were identified with cases in which different prices or different unit costs may be associated with the performance of different firms producing the same outputs and utilizing the same inputs. Types of price-cost efficiencies were therefore identified and related to extensions of the customary production possibility sets that reflected the unit price and cost differences.

Standard approaches were also extended in models that permit substitutions so that worsening of some inputs or outputs may be made in order to improve

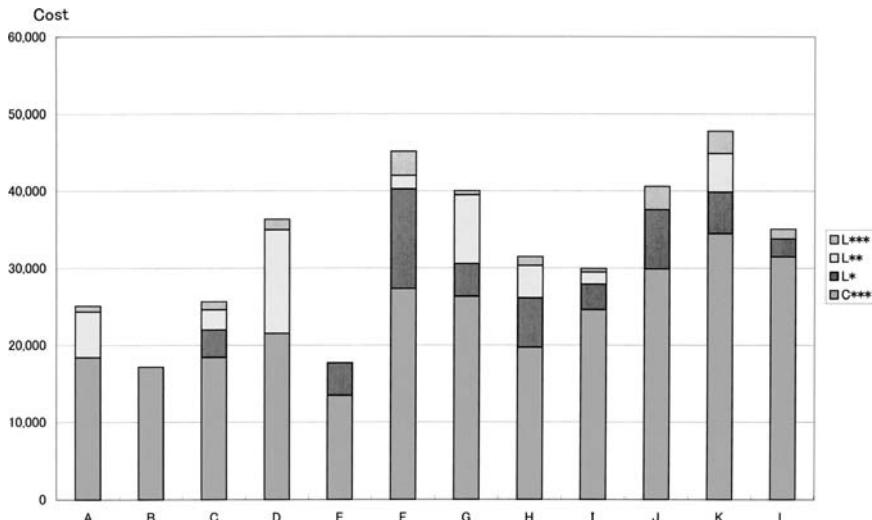


Figure 8.2. Decomposition of Actual Cost

other inputs or outputs. We also extended the traditional Farrell-Debreu cost efficiency measures and introduced new ones that can deal with non-identical cost and price situations. Furthermore, we provided a decomposition of the actual cost into the minimum cost and into losses due to other inefficiencies.

Problems in the use of these concepts may be encountered because many companies are unwilling to disclose their unit costs. As noted by Farrell, unit prices may also be subject to large fluctuations. One may, of course, use averages or other summaries of such prices and also deduce or infer which unit costs are applicable. However, this may not be satisfactory because in many cases accurate costs and prices may not really reflect criteria that are being used. Cases in point include attempts to evaluate public programs such as education, health, welfare, or military and police activities.

Earlier in this book a variety of approaches were suggested that can be applied to these types of problems. This includes the case of assurance regions and like concepts, as treated in Chapter 6, which can replace exact knowledge of prices and costs with corresponding bounds on their values. When this is done, however, the precise relations between allocative, overall and technical efficiencies may become blurred.

8.6 NOTES AND SELECTED BIBLIOGRAPHY

The concepts of cost efficiency related subjects were introduced by M.J. Farrell (1957) and G. Debreu (1951) and developed into implementable form by Färe, Grosskopf and Lovell (1985) using linear programming technologies. Cooper, Park and Pastor (1999)⁸ extended these treatments to the Additive models with a new “translation invariant” measure named “RAM.”

Inadequacies in attempts to move from “technical” to price based or cost based efficiencies were identified by Tone (2002). In response a new approach to cost efficiency was developed by Tone (2002) and further extended to decompositions of cost efficiency by Tone and Tautsui (2004) in a form they applied to Japan-US electric utilities comparisons. Tone and Sahoo (2005)⁹ applied the new cost efficiency model to examine the performance of Life Insurance Corporation (LIC) of India and found a significant heterogeneity in the cost efficiency scores over the course of 19 years. See also Tone and Sahoo (2005)¹⁰ in which the issues of cost elasticity are extensively discussed based on the new cost efficiency model. Fukuyama and Weber (2004)¹¹ developed a variant of the new cost efficiency model using “directional distance functions” introduced in Chambers, Chung and Färe (1996)¹² to measure inefficiency. See Färe, Grosskopf and Whittaker (2004)¹³ for an updated survey. See, however, Ray (2004),¹⁴ p.95, who identified a deficiency in the failure of “directional distance functions” to account for nonzero slacks in the measures of efficiency.

8.7 RELATED DEA-SOLVER MODELS FOR CHAPTER 8

(New-)Cost-C(V) Cost-C(V) code evaluates the *cost efficiency* of each DMU as follows. First we solve the LP problem below:

$$\begin{aligned} \min \quad & \sum_{i=1}^m c_i x_i \\ \text{subject to} \quad & x_i \geq \sum_{j=1}^n x_{ij} \lambda_j \quad (i = 1, \dots, m) \\ & y_{ro} \leq \sum_{i=1}^n y_{ri} \lambda_i \quad (r = 1, \dots, s) \\ & L \leq \sum_{j=1}^n \lambda_j \leq U \\ & \lambda_j \geq 0 \quad \forall j, \end{aligned} \tag{8.56}$$

where c_i is the unit cost of the input i . This model allows substitutions in inputs. Based on an optimal solution $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ of the above LP, the cost efficiency of DMU_o is defined as

$$E_C = \frac{\mathbf{c}\mathbf{x}^*}{\mathbf{c}\mathbf{x}_o}. \tag{8.57}$$

The code “Cost-C” solves the case $L = 0, U = \infty$ (the case of constant returns to scale) and “Cost-V” for the case $L = U = 1$ (the variable returns to scale case).

The data set (X, Y, C) should be prepared in an Excel Workbook under an appropriate Worksheet name, e.g., “.Data”, prior to execution of this code. See the sample format displayed in Figure B.6 in Section B.5 of Appendix B and refer to explanations above the figure.

The results will be obtained in the Worksheets of the selected Workbook: “Score”, “Projection” (projection onto the efficient frontier), “Graph1”, “Graph2” and “Summary.”

New-Cost-C(V) solves the model described in [NCost] (8.21). Data format is the same with the Cost-C(V) model.

(New-)Revenue-C(V) Revenue-C(V) code solves the following revenue maximization program for each DMU:

$$\begin{aligned} \max \quad & \mathbf{p} \mathbf{y} \\ \text{subject to} \quad & \mathbf{x}_o \geq X \boldsymbol{\lambda} \\ & \mathbf{y} \leq Y \boldsymbol{\lambda} \\ & L \leq e \boldsymbol{\lambda} \leq U \\ & \boldsymbol{\lambda} \geq 0, \end{aligned}$$

where the vector $\mathbf{p} = (p_1, \dots, p_s)$ expresses the unit prices of the output. This model allows substitutions in outputs.

The code “Revenue-C” solves the case $L = 0$. $U = \infty$ (the case of constant returns to scale) and “Revenue-V”, the case $L = U = 1$ (the variable returns to scale case).

Based on an optimal solution \mathbf{y}^* of this program, the *revenue efficiency* is defined as

$$E_R = \frac{\mathbf{p} \mathbf{y}_o}{\mathbf{p} \mathbf{y}^*}. \quad (8.58)$$

E_R satisfies $0 < E_R \leq 1$, provided $\mathbf{p} \mathbf{y}_o > 0$. See the sample data format displayed in Figure B.7 in Section B.5 of Appendix B and refer to the explanations above the figure.

The results will be obtained in the Worksheets of the selected Workbook: “Score”, “Projection” (projection onto the efficient frontier), “Graph1,” “Graph2,” and “Summary.”

New-Revenue-C(V) solves the model described in [NRevenue] (8.27). Data format is the same with Revenue-C(V).

(New-)Profit-C(V) Profit-C(V) code solves the LP problem defined in (8.8) for each DMU. Based on an optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$, the *profit efficiency* is defined as

$$E_P = \frac{\mathbf{p} \mathbf{y}_o - c \mathbf{x}_o}{\mathbf{p} \mathbf{y}^* - c \mathbf{x}^*}. \quad (8.59)$$

Under the assumption $\mathbf{p} \mathbf{y}_o > c \mathbf{x}_o$, we have $0 < E_P \leq 1$ and DMU_o is *profit efficient* if $E_P = 1$.

The data format is a combination of *Cost* and *Revenue* models. The cost columns are headed by (C) for input names and the price column are headed by (P) for output names.

The results will be obtained in the Worksheets of the selected Workbook: “Score”, “Projection” (projection onto the efficient frontier), “Graph1,” “Graph2,” and “Summary.”

New-Profit-C(V) solves the model described in [NProfit] (8.29). Data format is the same with Profit-C(V).

Ratio-C(V) This code solves the LP problem defined in (8.31). Based on the optimal solution (x^*, y^*) , the *ratio* (revenue/cost) *efficiency* is defined as

$$E_{RC} = \frac{p_o y_o / c_o x_o}{p_o y^* / c_o x^*},$$

which satisfies $0 < E_{RC} \leq 1$ and DMU_o is *ratio efficient* if $E_{RC} = 1$.

The data format is a combination of *Cost* and *Revenue* models. The cost columns are headed by (C) for input names and the price column are headed by (P) for output names.

The results will be obtained in the Worksheets of the selected Workbook: “Score”, “Projection” (projection onto the efficient frontier), “Graph1,” “Graph2,” and “Summary.”

8.8 PROBLEM SUPPLEMENT FOR CHAPTER 8

Problem 8.1

In Section 8.4, the actual cost is decomposed into the *sum* of the minimum cost and losses due to technical, price and allocative inefficiencies.

Can you decompose it in the *productive* form (not in the *sum* form)?

Suggested Response : Define the following efficiency measures:

- C^{***}/C = cost efficiency (CE)
- C^*/C = technical efficiency (TE)
- C^{**}/C^* = price efficiency (PE)
- C^{***}/C^{**} = allocative efficiency (AE)

Then, we have:

$$CE = TE \times PE \times AE.$$

Problem 8.2

Write out the decomposition of the actual profit in the same vein as described in Section 8.4.

Suggested Response : The actual profit of DMU (x_o, y_o) is calculated as:

$$E_o = p_o y_o - c_o x_o. \quad (8.60)$$

Using radial or non-radial technical efficiency models, e.g., the CCR and SBM, we project the DMU onto the efficient frontier and obtain the technically efficient $(\bar{x}_o^*, \bar{y}_o^*)$ with profit given by

$$E_o^* = p_o \bar{y}_o^* - c_o \bar{x}_o^* (\geq p_o y_o - c_o x_o = E_o). \quad (8.61)$$

Thus, the *loss due to technical inefficiency* is evaluated as

$$L_o^* = E_o^* - E_o. \quad (8.62)$$

We formulate the new cost-price based production possibility set as

$$P_{cp} = \{(\bar{x}, \bar{y}) | \bar{x} \geq \bar{X}\lambda, \bar{y} \leq \bar{Y}\lambda, \lambda \geq 0\}, \quad (8.63)$$

where $\bar{X} = (\bar{x}_1, \dots, \bar{x}_n)$ with $\bar{x}_j = (c_{1j}x_{1j}^*, \dots, c_{mj}x_{mj}^*)$ and $\bar{Y} = (\bar{y}_1, \dots, \bar{y}_n)$ with $\bar{y}_j = (p_{1j}y_{1j}^*, \dots, p_{sj}y_{sj}^*)$. On this PPS we form a technical efficiency model similar to [NTech-2] (8.40) as follows:

$$\begin{aligned} [\text{NTech-3}] \quad & \rho^* = \min_{\bar{x}, \bar{y}, \rho, \mu} \rho \\ \text{subject to} \quad & \rho \bar{x}_o \geq \bar{X}\mu = \bar{x} \\ & \bar{y}_o \leq \bar{Y}\mu = \bar{y} \\ & \mu \geq 0. \end{aligned} \quad (8.64)$$

Let the optimal solution of [NTech-3] be $(\bar{x}_o^*, \bar{y}_o^*, \mu^*, \rho^*)$. (Note that, instead of [NTech-3], we can apply the non-radial and non-oriented SBM for obtaining $(\bar{x}_o^*, \bar{y}_o^*)$.)

We then have the technical and cost-price efficient profit given by

$$E_o^{**} = e \bar{y}_o^* - e \bar{x}_o^* (\geq E_o^*). \quad (8.65)$$

The loss due to cost-price inefficiency is estimated by

$$L_o^{**} = E_o^{**} - E_o^* (\geq 0). \quad (8.66)$$

Lastly, we solve the following max profit model on P_{cp} .

$$\begin{aligned} [\text{NProfit-2}] \quad & e \bar{y}_o^{**} - e \bar{x}_o^{**} = \max_{\bar{x}, \bar{y}, \lambda} e \bar{y} - e \bar{x} \\ \text{subject to} \quad & \bar{x}_o \geq \bar{X}\lambda = \bar{x} \\ & \bar{y}_o \leq \bar{Y}\lambda = \bar{y} \\ & \lambda \geq 0. \end{aligned} \quad (8.67)$$

Let the optimal solution be $(\bar{x}_o^{**}, \bar{y}_o^{**})$. The allocative efficient profit is given by

$$E_o^{***} = e \bar{y}_o^{**} - e \bar{x}_o^{**}. \quad (8.68)$$

The loss due to allocative inefficiency is evaluated by

$$L_o^{***} = E_o^{***} - E_o^{**} (\geq 0). \quad (8.69)$$

Summing up, we have the decomposition of the actual profit into the maximum profit and the losses due to technical, cost-price and allocative inefficiencies as follows:

$$E_o = E_o^{***} - L_o^* - L_o^{**} - L_o^{***}. \quad (8.70)$$

Problem 8.3

Most of the models in this Chapter are developed under the constant returns-to-scale (CRS) assumption. Can you develop them under the variable returns-to-scale assumption?

Suggested Response : This can be done by adding the convex constraint on the intensity variable λ as follows:

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

Problem 8.4

In connection with the preceding problem, can you incorporate the scale inefficiency effect, i.e., loss due to scale inefficiency, in the model?

Suggested Response : In the cost efficiency case, we first solve the input-oriented technical efficiency model under the VRS assumption. Let the optimal solution be $(\mathbf{x}_o^{VRS}, \mathbf{y}_o^{VRS})$. The cost and loss due to this projection are, respectively:

$$C_o^{VRS} = \mathbf{c}_o \mathbf{x}_o^{VRS}, \quad L_o^{VRS} = C_o - C_o^{VRS}, \quad (8.71)$$

where C_o is the actual observed cost of DMU $(\mathbf{x}_o, \mathbf{y}_o)$. Then we construct the data set (X^{VRS}, Y^{VRS}) consisting of $(\mathbf{x}_j^{VRS}, \mathbf{y}_j^{VRS})$ $j = 1, \dots, n$. We next evaluate the technical efficiency of $(\mathbf{x}_o^{VRS}, \mathbf{y}_o^{VRS})$ with respect to (X^{VRS}, Y^{VRS}) under the constant returns-to-scale (CRS) assumption. Let the optimal solution be $(\mathbf{x}_o^*, \mathbf{y}_o^*)$, with its cost $C_o^* = \mathbf{c}_o \mathbf{x}_o^*$ ($\leq C_o^{VRS}$). Thus, we obtain the loss due to scale inefficiency as follows:

$$L_o^{Scale} = C_o^{VRS} - C_o^*. \quad (8.72)$$

Referring to (8.52)-(8.55), we can decompose the actual cost into four losses with minimum cost as follows:

$$L_o^{VRS} = C_o - C_o^{VRS} (\geq 0) \text{ Loss due to Pure Tech. Inefficiency} \quad (8.73)$$

$$L_o^{Scale} = C_o^{VRS} - C_o^* (\geq 0) \text{ Loss due to Scale Inefficiency} \quad (8.74)$$

$$L_o^{**} = C_o^* - C_o^{**} (\geq 0) \text{ Loss due to Price Inefficiency} \quad (8.75)$$

$$L_o^{***} = C_o^{**} - C_o^{***} (\geq 0) \text{ Loss due to Allocative Inefficiency} \quad (8.76)$$

$$C_o = L_o^{VRS} + L_o^{Scale} + L_o^{**} + L_o^{***} + C_o^{***}. \quad (8.77)$$

Problem 8.5

The concluding part of Section 5.5, Chapter 5, quoted Dr. Harold Wein, a steel industry consultant, who believed that the concept of returns to scale, as formulated in economics, is useless because increases in plant size are generally accompanied by mix changes in outputs or inputs — or both.

Assignment : Formulate the responses an economist might make to this criticism.

Suggested Response : Under the assumption of profit maximization, as employed in economics, both scale and mix are determined simultaneously. This is consistent with Dr. Wein's observation from his steel industry experience.

One response to Dr. Wein's criticism of the scale concept as employed in economics (which holds mix constant) is to note that economists are interested in being able to distinguish between scale and mix changes when treating empirical data *after* the decisions have been made. Economists like Farrell¹⁵ and Debreu¹⁶ contributed concepts and methods for doing this, as cited in this chapter (and elsewhere in this text), and these have been further extended by economists like Färe, Grosskopf and Lovell whose works have also been cited at numerous points in this text.¹⁷

The above response is directed to *ex post* analyses. Another response is directed to whether economics can contribute to these decisions in an *ex ante* fashion. The answer is that there is a long-standing literature which provides guidance to scale decisions by relating marginal (incremental) costs to marginal (incremental) receipts. Generally speaking, returns to scale will be increasing as long as marginal costs are below average (unit) costs. The reverse situation applies when returns to scale is decreasing. Equating marginal costs to marginal receipts will determine the best (most profitable) scale size.

This will generally move matters into the region of decreasing returns to scale. Reasons for this involve issues of stability of solutions which we cannot treat here. Marginal cost lying below average unit cost in regions of increasing returns to scale means that average unit cost can be decreased by incrementing outputs. Hence if marginal receipts equal or exceed average unit cost it is possible to increase total profit by incrementing production.

The above case refers to single output situations. Modifications are needed to allow for complementary and substitution interactions when multiple outputs are involved. This has been accomplished in ways that have long been available which show that the above rule continues to supply general guidance. Indeed, recent years have seen this extended to ways for determining "economics of scope" in order to decide whether to add or delete product lines while simultaneously treating mix and scale decisions. See W.S. Baumol, J.C. Panzar and R.D. Willig (1982) *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt Brace Jovanovich).

Much remains to be done in giving the above concepts implementable form — especially when technical inefficiencies are involved and errors and uncertainties are also involved. A start has been made in the form of what are referred

to as “stochastic frontier” approaches to statistical estimation. However, these approaches are, by and large, confined to the use of single output regressions.

Comment : As already noted, profit maximization requires a simultaneous determination of the best (i.e., most profitable) combination of scale, scope, technical and mix efficiencies. The models and methods described in earlier chapters can then be used to determine what was done and whether and where any inefficiencies occurred.

Problem 8.6

Prove that the *profit ratio* model in (8.31) does not suffer from the inadequacies pointed out in Section 8.3.

Suggested Response : Since the profit ratio efficiency is defined as ratio of ratios (between revenue and cost), it is invariant when we double both the unit cost and price.

Notice that the traditional profit efficiency model [Profit] (8.8) gives the same efficiency value when we double both the unit cost and price. This is unacceptable.

Notes

1. M.J. Farrell (1957), “The Measurement of Productive Efficiency,” *Journal of the Royal Statistical Society Series A*, 120, III, pp.253-281. G. Debreu (1951), “The Coefficient of Resource Utilization,” *Econometrica* 19, pp.273-292.
2. R. Färe, S. Grosskopf and C.A.K. Lovell, *Measurement of Efficiency of Production* (Boston: Kluwer-Nijhoff Publishing Co., Inc., 1985).
3. This means that P lies below the production possibility surface and hence is inefficient.
4. K. Tone (2002), “A Strange Case of the Cost and Allocative Efficiencies in DEA,” *Journal of the Operational Research Society* 53, pp.1225-1231.
5. This ratio model was introduced in K. Tone (1993), *Data Envelopment Analysis* (in Japanese)(Tokyo: JUSE Press, Ltd.).
6. W.W. Cooper, Z. Huang, S. Li and J.T. Pastor (2005) “Aggregation with Enhanced Russell Measure in DEA,” *European Journal of Operational Research* (forthcoming).
7. K. Tone and M. Tsutsui (2004), “Decomposition of Cost Efficiency and its Application to Japan-US Electric Utility Comparisons,” Research Report Series I-2004-0004, GRIPS (National Graduate Institute for Policy Studies), also forthcoming in *Socio-Economic Planning Sciences*.
8. W.W. Cooper, K.S. Park and J.T. Pastor (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, pp.5-42.
9. K. Tone and B.K. Sahoo (2005), “Evaluating Cost Efficiency and Returns to Scale in the Life Insurance Corporation of India Using Data Envelopment Analysis,” *Socio-Economic Planning Sciences* 39, pp.261-285.
10. K. Tone and B.K. Sahoo (2005), “Cost-Elasticity: A Re-Examination in DEA,” *Annals of Operations Research* (forthcoming).
11. H. Fukuyama and W.L. Weber (2004), “Economic Inefficiency Measurement of Input Spending When Decision-making Units Face Different Input Prices,” *Journal of the Operational Research Society* 55, pp.1102-1110.
12. R.G. Chambers, Y. Chung and R. Färe (1996) “Benefit and Distance Functions,” *Journal of Economic Theory* 70, pp.407-418.

13. R. Färe, S. Grosskopf and G. Whittaker (2004) "Distance Functions," Chapter 5 in W.W. Cooper, L.M. Seiford and J. Zhu, eds., *Handbook on Data Envelopment Analysis* (Norwell Mass., Kluwer Academic Publishers).
14. S. Ray (2004) *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research* (Cambridge University Press).
15. See Note 1. See also M.J. Farrell and M. Fieldhouse (1962), "Estimating Efficient Production Functions Under Increasing Returns to Scale," *Journal of the Royal Statistical Society Series A*, 125, Part 2, pp.252-267.
16. See the Note 1 reference.
17. See the Note 2 above. See also R. Färe, S. Grosskopf and C.A.K. Lovell (1994), *Production Frontiers* (Cambridge: Cambridge University Press).

9 DATA VARIATIONS

9.1 INTRODUCTION

We have now covered a wide variety of topics ranging from straightforward DEA models and their uses and extending to modifications such as are incorporated in assurance regions or in the treatment of variables with values that have been exogenously fixed to varying degrees. This, however, does not end the possibilities. New uses of DEA with accompanying new developments and extensions continue to appear.

An attempt to cover all of these topics would be beyond the scope of this text — and perhaps be impossible of achievement as still more uses and new developments continue to appear. Hence the course we follow is to provide relatively brief introductions to some of these topics. This will include sensitivity analysis, stochastic-statistical characterizations and probabilistic formulations with their associated approaches.

9.2 SENSITIVITY ANALYSIS

9.2.1 Degrees of Freedom

The topic of sensitivity (= stability or robustness) analysis has taken a variety of forms in the DEA literature. One part of this literature studies the responses when DMUs are deleted or added to the set being considered or when outputs or inputs are added or withdrawn from consideration. See Wilson (1995).¹ See

also the discussion of “window analysis” later in this chapter. Another part of this literature deals with the increases or decreases in the number of inputs and outputs to be treated. Analytically oriented treatments of these topics are not lacking² but most of this literature has taken the form of simulation studies, as in Banker *et al.* (1996).³

Comment : As in statistics or other empirically oriented methodologies, there is a problem involving degrees of freedom, which is compounded in DEA because of its orientation to *relative* efficiency. In the envelopment model, the number of degrees of freedom will increase with the number of DMUs and decrease with the number of inputs and outputs. A rough rule of thumb which can provide guidance is as follows.

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

where n = number of DMUs, m = number of inputs and s = number of outputs.

9.2.2 Algorithmic Approaches

Attention to this topic of sensitivity analysis in DEA was initiated in Charnes *et al.* (1985)⁴ which built on the earlier work in Charnes and Cooper (1968)⁵ after noting that variations in the data for the DMU_o under evaluation could alter the inverse matrix used to generate solutions in the usual simplex algorithm computer codes. (See expressions (3.52)-(3.53) in Chapter 3.) Proceeding along the path opened by the latter publication (by Charnes and Cooper) this work is directed to the use of algorithms that avoid the need for additional matrix inversions. Originally confined to treating a single input or output this line of work was extended and improved in a series of papers published by Charnes and Neralic.⁶

We do not pursue these algorithmic approaches here. We turn instead to other approaches where attention is confined to transitions from efficient to inefficient status.

9.2.3 Metric Approaches

Another avenue for sensitivity analysis opened by Charnes *et al.* (1992)⁷ bypasses the need for these kinds of algorithmic forays by turning to metric concepts. The basic idea is to use concepts such as “distance” or “length” (= norm of a vector) in order to determine “radii of stability” within which the occurrence of data variations will not alter a DMU’s classification from efficient to inefficient status (or *vice versa*).

The resulting classifications can range from “unstable” to “stable” with the latter identified by a radius of some finite value within which no reclassification will occur. Points like E or F in Figure 9.1 provide examples identified as stable. A point like A , however, is unstable because an infinitesimal perturbation to the left of its present position would alter its status from inefficient to efficient.

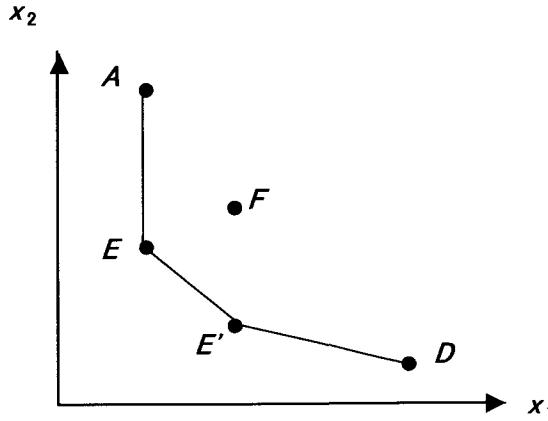


Figure 9.1. Stable and Unstable DMUs

A variety of metrics and models are examined but here attention will be confined to the Chebychev ($= l_\infty$) norm, as in the following model taken from Charnes, Haag, *et al.* (1992, p.795),⁷

$$\begin{aligned} & \max \delta && (9.1) \\ \text{subject to} \quad & y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ - \delta d_r^+, \quad r = 1, \dots, s \\ & x_{io} = \sum_{j=1}^n x_{ij} \lambda_j + s_i^- + \delta d_i^-, \quad i = 1, \dots, m \\ & 1 = \sum_{j=1}^n \lambda_j \end{aligned}$$

with all variables (including δ) constrained to be nonnegative while the d_r^+ and d_i^- are fixed constants (to serve as weights) which we now set to unity.

With all $d_i^- = d_r^+ = 1$ the solution to (9.1) may be written

$$\begin{aligned} & \sum_{j=1}^n y_{rj} \lambda_j^* - s_r^{+*} = y_{ro} + \delta^*, \quad r = 1, \dots, s && (9.2) \\ & \sum_{j=1}^n x_{ij} \lambda_j^* + s_i^{-*} = x_{io} - \delta^*, \quad i = 1, \dots, m \end{aligned}$$

This shows that we are improving all outputs and inputs to the maximum that this metric allows consistent with the solution on the left.

The formulation in (9.2) is for an inefficient DMU which continues to be inefficient for all data alteration from y_{ro} to $y_{ro} + \delta^*$ and from x_{io} to $x_{io} - \delta^*$. This is interpreted to mean that no reclassification to efficient status will occur within the open set defined by the value of $0 \leq \delta^*$ — which is referred to as

a “radius of stability.” See, for example, the point F in Figure 9.2 which is centered in the square (or box) defined by this C (=Chebyshev) norm which is referred to as a “unit ball.”⁸

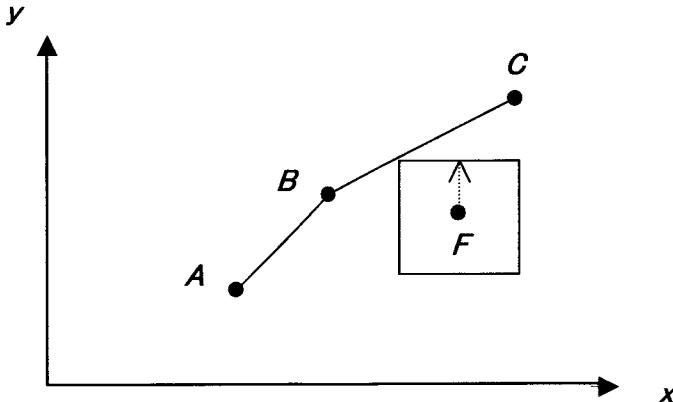


Figure 9.2. A Radius of Stability

The above model dealt with improvements in both inputs and outputs that could occur for an inefficient point before its status would change to efficient — as in the upper left hand corner of the box surrounding F in Figure 9.2. The treatment of efficient points proceeds in the direction of “worsening” outputs and inputs as in the following model.

$$\begin{aligned} & \min \delta && (9.3) \\ \text{subject to} \quad & y_{ro} = \sum_{j=1, j \neq o}^n y_{rj} \lambda_j - s_r^+ - \delta, \quad r = 1, \dots, s \\ & x_{io} = \sum_{j=1, j \neq o}^n x_{ij} \lambda_j + s_i^- + \delta, \quad i = 1, \dots, m \\ & 1 = \sum_{j=1, j \neq o}^n \lambda_j \end{aligned}$$

where, again, all variables are constrained to be nonnegative.

In this case $j \neq o$ refers to the efficient DMU _{o} that is being analyzed. Otherwise, as in the following definition, the result will always be unstable.⁹

Definition 9.1 *The coordinates of the point associated with an efficient DMU will always have both efficient and inefficient points within a radius of $\varepsilon > 0$ however small the value of ε .*

Definition 9.2 *Any point with the above property is unstable.*

To see that this property is not confined to points associated with efficient DMUs, note that A in Figure 9.1 has this property since a slight variation to the left will change its status from inefficient to efficient. In any case, a solution, δ^* , provides a radius in the Chebychev norm that is to be attained before an efficient DMU is changed to inefficient status.

To see what is happening in the case of an efficient point refer to B in Figure 9.2. The radius of stability for this point would be determined by the “worsenings” allowed in (9.3) until the line connecting A and C is reached. This follows from the fact that worsenings which only move B to a new point which is on the left of this line will not affect its efficiency status.

Comment : The above formulations are recommended only as the *start* for a sensitivity analysis by Charnes *et al.*¹⁰ because, *inter alia*, this norm does not reflect any nonzero slacks which may be present.¹¹ It might be supposed that omitting the DMU to be evaluated from the right-hand side of (9.3) could lead to non-solution possibilities. This is not the case. Solutions to (9.3) always exist, as is proved in W.W. Cooper, S. Li, L.M. Seiford, K. Tone, R.M. Thrall and J. Zhu (2001) “Sensitivity and Stability Analysis in DEA: Some Recent Developments,” *Journal of Productivity Analysis* 15, pp.217-246. See also L.M. Seiford and J. Zhu (1998) “Sensitivity Analysis of DEA Models for Simultaneous Changes in All of the Data,” *Journal of the Operational Research Society* 49, pp.1060-1071 as well as Seiford and Zhu (1999) “Infeasibility of Super-Efficiency Data Envelopment Analysis Models,” *INFOR* 37, pp.174-187.

9.2.4 Multiplier Model Approaches

The above approaches treat one DMU at a time. However, this needs to be extended for treating cases where the DMUs are numerous and it is not clear which ones require attention. Ideally it should be possible to vary all data simultaneously until the status of at least one DMU is changed from inefficient to efficient or *vice versa*. A third approach initiated by R.G. Thompson and R.M. Thrall¹² and their associates moves in this direction in a manner that we now describe.

For this purpose we record the following dual pair from Thompson *et al.* (1996).¹³

Envelopment Model	Multiplier Model	(9.4)
minimize _{θ, λ} θ	maximize _{u, v} $z = uy_o$	
subject to	subject to	
$Y\lambda \geq y_o$	$u \geq 0$	
$\theta x_o - X\lambda \geq 0$	$v \geq 0$	
$\lambda \geq 0$	$uY - vX \leq 0$	
θ unrestricted	$vx_o = 1,$	

where Y, X and y_o, x_o are data matrices and vectors of outputs and inputs, respectively, and λ, u, v are vectors of variables (λ : a column vector, u and v : row vectors). θ , a scalar, which can be positive, negative or zero in the envelopment model is the source of the condition $vx_o = 1$ in the multiplier model.

No allowance for nonzero slacks is made in the objective of the above envelopment model. Hence the variables in the multiplier model are constrained only to be nonnegative. That is, the positivity requirement associated with the non-Archimedean element, ε , is absent from both members of this dual pair. Thompson *et al.* refer to Charnes, Cooper and Thrall (1991)¹⁴ to justify this omission of non-Archimedean elements. For present purposes, however, we note only that these sensitivity analyses are centered around the set, E , of efficient extreme points and these points always have a unique optimum with nonzero slacks.

We also note that the analysis is carried forward via the multiplier models¹⁵ by Thompson, *et al.* This makes it possible to exploit the fact that the values u^*, v^* which are optimal for the DMU being evaluated will remain valid over some (generally positive) range of variation in the data.¹⁶

Following Thompson, *et al.* we try to exploit this property by defining a new vector $w = (u, v)$ and a function $h_j(w)$ as follows

$$h_j(w) = \frac{f_j(w)}{g_j(w)} = \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}}. \quad (9.5)$$

Next, let

$$h_o(w) = \max_{j=1, \dots, n} h_j(w) \quad (9.6)$$

so that

$$h_o(w) \geq h_j(w) \quad \forall j. \quad (9.7)$$

It is now to be noted that (9.5) returns matters to the CCR ratio form which was introduced as early as Chapter 2, Section 2.2. Hence we need not be concerned with continued satisfaction of the norm condition, $vx_o = 1$ in (9.4), as we study variations in the data.

When an *optimal* w^* does not satisfy (9.7), the DMU_o being evaluated is said to be “radial inefficient.” The term is appropriate because this means that $\theta^* < 1$ will occur in the envelopment model. The full panoply of relations between the CCR ratio, multiplier and envelopment models is thus brought into play without any need for extensive computations or analyses.

Among the frontier points for which $\theta^* = 1$, attention is directed by Thompson *et al.* to “extreme efficient points” — i.e., points in the set E which, for some multiplier w^* ,

$$h_o(w^*) > h_j(w^*) \quad \forall j \neq o. \quad (9.8)$$

This (strict) inequality will generally remain valid over some range of variation in the data. Hence, in more detail we will have

$$h_o(w^*) = \frac{\sum_{r=1}^s u_r^* y_{ro}}{\sum_{i=1}^m v_i^* x_{io}} > \frac{\sum_{r=1}^s u_r^* y_{rj}}{\sum_{i=1}^m v_i^* x_{ij}} = h_j(w^*) \quad \forall j \neq o, \quad (9.9)$$

which means that DMU_o is more efficient than any other DMU_j and hence will be rated as fully efficient by DEA.

Thompson, *et al.* employ a ranking principle which they formulated as: “If DMU_o is more efficient than any other DMU_j relative to the vector w^* , then DMU_o is said to be top ranked.” Holding w^* fixed, the data are then varied and DMU_o is said to be “top ranked” if (9.8) continues to hold.

Thompson, *et al.* allow the data to vary in several different ways which include allowing the data variations to occur at random. Among these possibilities we examine only the following one. For DMU_o , which is extreme efficient, the outputs are all decreased and the inputs are all increased by a stipulated amount (or percentage). This same treatment is accorded to the other DMUs which are efficient. For the other DMU_j , the reverse adjustment is made: All outputs are increased and all inputs are decreased in these same amounts (or percentages). In this way the ratio in (9.8) for DMU_o will be decreased along with the other extreme efficient DMUs while the ratios for the other DMU_j will be increased. Continuing in this manner a reversal can be expected to occur at some point — in which case DMU_o will no longer be “top ranked” and it will then lose the status of being fully (DEA) efficient.

Table 9.1 taken from Thompson *et al.* (1994) will be used to illustrate the procedure in a simple manner by varying only the data for the inputs x_1, x_2 in this table. To start this sensitivity analysis, Table 9.2 records the initial

Table 9.1. Data for a Sensitivity Analysis

	E-Efficient*			Not Efficient		
DMU	1	2	3	4	5	6
Output: y	1	1	1	1	1	1
Input: x_1	4	2	1	2	3	4
Input: x_2	1	2	4	3	2	4

* E-Efficient = Extreme Point Efficient

solutions by applying the multiplier model for (9.8) to these data for each of DMU_1, DMU_2 and DMU_3 which are all extreme point efficient.¹⁷ As can be

Table 9.2. Initial Solutions

DMU	DMU ₁	DMU ₂	DMU ₃
	$h_j(w^1)$	$h_j(w^2)$	$h_j(w^3)$
1	1.000	0.800	0.400
2	0.714	1.000	0.714
3	0.400	0.800	1.000
4	0.500	0.800	0.667
5	0.667	0.800	0.550
6	0.357	0.500	0.357

seen these solutions show DMU₁, DMU₂ and DMU₃ to be top ranked in their respective columns.

The gaps between the top and other ranks from these results show that some range of data variation can be undertaken without changing this top-ranked status. To start we therefore introduce 5% increases in each of x_1 and x_2 for DMU₁, DMU₂ and DMU₃ and thereby worsen their performances. Simultaneously we decrease these inputs by 5% for the other DMUs to obtain Table 9.3.

Table 9.3. Results of 5% Data Variations

DMU	DMU ₁	DMU ₂	DMU ₃
	$h_j(w^1)$	$h_j(w^2)$	$h_j(w^3)$
1	0.952	0.762	0.381
2	0.680	0.952	0.680
3	0.381	0.762	0.952
4	0.526	0.842	0.702
5	0.702	0.842	0.526
6	0.376	0.526	0.376

The values of the $h_j(\mathbf{w})$ resulting from these data variations are portrayed in Table 9.3. As can be seen, each of DMU₁, DMU₂ and DMU₃ maintain their “top ranked status” and hence continue to be DEA fully efficient (relatively). Nor is this the end of the line. Continuing in this 5% increment-decrement fashion, as Thompson, *et al.* (1994) report, a 15% increment-decrement is needed for a first displacement in which DMU₂ is replaced by DMU₄ and DMU₅. Continuing further, a 20% increment-decrement is needed to replace DMU₁ with DMU₅ and, finally, still further incrementing and decrementing is needed to

replace DMU₃ with DMU₄ as top ranked.

Comment : Note that the $h_j(\mathbf{w})$ values for all of the efficient DMUs decrease in every column when going from Table 9.2 to Table 9.5 and, simultaneously, the $h_j(\mathbf{w})$ values increase for the inefficient DMUs. The same behavior occurs for the other data variations, including the random choices of data changes, used by Thompson, Thrall and their associates in other studies. As noted on page 401 of Thompson *et al.* (1994) this robust behavior is obtained for extreme efficient DMUs which are identified by their satisfaction of the Strong Complementary Slackness Condition (described in Section A.8 of our Appendix A) for which a gap will appear like ones between the top and second rank shown in every column of Table 9.2. In fact, the choice of \mathbf{w}^* can affect the degree of robustness as reported in Thompson *et al.* (1996) where use of an interior point algorithm produces a \mathbf{w}^* closer to the analytic center and this considerably increases the degree of robustness for the above example. For a more detailed treatment see Cooper, Li, Seiford and Zhu.¹⁸

9.3 STATISTICAL APPROACHES

Treatment of data variations by statistical methods has taken a variety of forms in DEA and related literatures. More precisely, Banker (1993)¹⁹ and Banker and Natarasan (2004)²⁰ show that DEA provides a consistent estimator of arbitrary monotone and concave production functions when the (one-sided) deviations from such a production function are regarded as stochastic variations in technical inefficiency.²¹ Convergence is slow, however, since, as is shown by Korostolev *et al.* (1995),²² the DEA likelihood estimator in the single output - m input case converges at the rate $n^{-2/(1+m)}$ and no other estimator can converge at a faster rate.²³

The above approaches treat only the single output - multiple input case. Simar and Wilson (1998)²⁴ turn to “bootstrap methods” which enable them to deal with the case of multiple outputs and inputs. In this manner, the sensitivity of θ^* , the efficiency score obtained from the BCC model, can be tested by repeatedly sampling from the original samples. A sampling distribution of θ^* values is then obtained from which confidence intervals may be derived and statistical tests of significance developed.

All of this work represents significant new developments. More remains to be done, however, since neither Banker nor Simar and Wilson make any mention of how to treat nonzero slacks. Thus, it is not even clear that they are estimating efficiency frontiers.

Another line of research proceeds through what are referred to as “stochastic frontier regressions.” This line of work has a longer history which can (in a sense) be traced all the way back to Farrell (1957).²⁵ Subsequently extended by Aigner and Chu (1968)²⁶ this approach was given its first statistical formulation in Aigner, Lovell and Schmidt (1977) in a form that is now called the “composed error” approach.²⁷

To see what is involved in this “composed error” approach we start with the usual formulation of a statistical regression model as in

$$y = f(\mathbf{x}) + \varepsilon. \quad (9.10)$$

Here $f(\mathbf{x})$ is a prescribed (known) function with parameters to be estimated and ε represents random errors which occur in the dependent (regressand) variable, a scalar, and not in the independent (regressor) variables represented by the vector \mathbf{x} . The components of \mathbf{x} , we emphasize, are assumed to be known without error.

The concept of a “composed error” is represented by replacing ε with a 2-component term which we can represent as

$$\varepsilon = \nu - \tau. \quad (9.11)$$

Here ν represents the random error component which may be positive, negative or zero while τ is restricted to nonnegative ranges that represent values of y that fail to achieve the efficient frontier. The term $\tau \geq 0$ is usually assumed to have statistical distributions, such as the exponential or half normal, which are confined to nonnegative ranges that represent inefficiencies.

Following Jondrow, Lovell, Materov and Schmidt (1982),²⁸ the estimates of technical efficiency are obtained from

$$\hat{\tau} = - \left[\mu_\tau - \sigma \frac{f^*(\mu_\tau/\sigma)}{F^*(-\mu_\tau/\sigma)} \right] \quad (9.12)$$

where

$$\mu_\tau = \frac{\sigma_\tau^2}{\sigma_\nu^2 + \sigma_\tau^2} \quad \text{and} \quad \sigma^2 = \frac{\sigma_\nu^2 \sigma_\tau^2}{\sigma_\nu^2 + \sigma_\tau^2}$$

and where $f^*(\cdot)$ and $F^*(\cdot)$ represent the standard normal density and cumulative normal distribution functions, respectively, with mean μ and variance σ^2 . The efficiency corresponding to specified values for the components of \mathbf{x} are then estimated from

$$0 \leq e^{-\hat{\tau}} \leq 1 \quad (9.13)$$

which is equal to unity when $\hat{\tau} = 0$ and becomes 0 as $\hat{\tau} \rightarrow \infty$.

To see how this measure of efficiency is to be used we employ (9.10) and (9.11) in the following simple (two-input) version of a log-linear (=Cobb-Douglas) production function

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} e^\varepsilon = \beta_0 x_1^{\beta_1} x_2^{\beta_2} e^{\nu - \tau} \quad (9.14)$$

so that

$$ye^\tau = \beta_0 x_1^{\beta_1} x_2^{\beta_2} e^\nu. \quad (9.15)$$

Hence $ye^\tau = \hat{y}$ with $\hat{y} \geq y$ is estimated stochastically with inefficiency embodied in an output shortfall and not in overuses of either input.

It is possible to view these stochastic frontier regressions as competing with DEA and to study them from this standpoint as is done in Gong and Sickles

(1990),²⁹ for example, who bring a simulation approach to this task. Carried to an extreme the two approaches, DEA vs. Stochastic Frontier Regressions, can be regarded as mutually exclusive — as in Schmidt (1985).³⁰ An alternative view is also possible in which the two approaches can be used in complementary fashion. Ferrier and Lovell (1990),³¹ for example, use the two approaches to cross-check each other. In this approach, the objective is to avoid what Charnes, Cooper and Sueyoshi (1988)³² refer to as “methodological bias” when large issues of policy are being addressed. Indeed, it is possible to go a step further and join the two approaches in a common effort as in the example we now discuss.³³

Arnold *et al.* (1994)³⁴ describe an experience in which the use of Cobb-Douglas regressions yielded unsatisfactory results in an attempt to use this kind of regression approach in a study conducted under legislative mandate to develop methods for evaluating the performances of public schools in Texas. Using this same functional form, however, and applying it to the same body of data, Arnold *et al.* reported satisfactory results from a two-stage DEA-regression approach which proceeded in the following manner: In stage one all of the 640 schools in this study were submitted to treatment by DEA. The original Cobb-Douglas form was then extended to incorporate these results in the form of “dummy variables.” In this approach a school which had been found to be DEA efficient was associated with a value of unity for the dummy variables. A school which had been found to be DEA inefficient was assigned a value of zero. The regression was then recalculated and found to yield very satisfactory results.

The above study was followed by a simulation experiment which we now review for added insight.³⁵ For this purpose we replaced (9.10) with

$$y = 0.75x_1^{0.65}x_2^{0.55}e^\varepsilon. \quad (9.16)$$

In short, the above expression is used to generate all observations with the Cobb-Douglas form having known parameter values

$$\begin{aligned} \beta_0 &= 0.75 \\ \beta_1 &= 0.65 \\ \beta_2 &= 0.55 \end{aligned} \quad (9.17)$$

and e^ε is used to generate random variables which are then used to adjust the thus generated y values to new values which contain these random terms. This procedure conforms to the assumptions of both SF (=Stochastic Frontier) and OLS (=Ordinary Least Squares) regression uses.

The input values for x_1 and x_2 in (9.16) are generated randomly, as a bias avoiding mechanism, and these values are inserted in (9.16) to provide the truly efficient values of y after which the values of \hat{y} defined in (9.15) are then generated in the previously described manner.

The inputs are then adjusted to new values

$$\hat{x}_1 = x_1 e^{\tau_1} \text{ and } \hat{x}_2 = x_2 e^{\tau_2} \text{ with } \tau_1, \tau_2 \geq 0 \quad (9.18)$$

where τ_1 and τ_2 represent input-specific technical inefficiencies drawn at random to yield the corresponding input inefficiencies embedded in \hat{x}_1 and \hat{x}_2 .

This procedure, we note, violates the SF assumption that all inefficiencies are impounded only in the regressand, y . See the discussion immediately following (9.15). It also violates OLS since these \hat{x}_1 , \hat{x}_2 are not the input amounts used to generate the y values. Nevertheless it reproduces a situation in which the observed y (or \hat{y}) will tend to be too low for these inputs. Finally, to complete the experimental design, a subset of the observations, chosen at random, used the original x_1 , x_2 values rather than the \hat{x}_1 , \hat{x}_2 generated as in (9.18). This was intended to conform to the assumption that some DMUs are wholly efficient and it also made it possible (a) to determine whether the first-stage DEA identified the efficient DMUs in an approximately correct manner as well as (b) to examine the effects of such an efficient subset on the derived statistical estimates.

Further details on the experimental design may be found in Bardhan *et al.* (1998).

We therefore now turn to the estimating relations which took a logarithmic form as follows,

$$\ln y = \ln \beta_0 + \beta_1 \ln \hat{x}_1 + \beta_2 \ln \hat{x}_2 \quad (9.19)$$

and

$$\ln y = \ln \beta_0 + \beta_1 \ln \hat{x}_1 + \beta_2 \ln \hat{x}_2 + \delta D + \delta_1 D \ln \hat{x}_1 + \delta_2 D \ln \hat{x}_2 + \varepsilon \quad (9.20)$$

Table 9.4. OLS Regression Estimates without Dummy Variables

Case 1: EXPONENTIAL distribution of input inefficiencies

Parameter Estimates	<i>Case A</i> $\sigma_\varepsilon^2 = 0.04$ (1)	<i>Case B</i> $\sigma_\varepsilon^2 = 0.0225$ (2)	<i>Case C</i> $\sigma_\varepsilon^2 = 0.01$ (3)	<i>Case D</i> $\sigma_\varepsilon^2 = 0.005$ (4)
β_0	1.30* (0.19)	1.58* (0.15)	1.40* (0.13)	1.43* (0.10)
β_1	0.46* (0.024)	0.43* (0.02)	0.45* (0.016)	0.46* (0.013)
β_2	0.48* (0.02)	0.47* (0.013)	0.47* (0.01)	0.46* (0.01)

The asterisk “*” denotes statistical significance at 0.05 significance level or better. Standard errors are shown in parentheses.

The values for σ_ε^2 shown in the top row of the table represent the true variances for the statistical error distributions.

Table 9.5. Stochastic Frontier Regression Estimates without Dummy VariablesCase 1: *EXPONENTIAL* distribution of input inefficiencies

Parameter Estimates	<i>Case A</i> $\sigma_e^2 = 0.04$ (1)	<i>Case B</i> $\sigma_e^2 = 0.0225$ (2)	<i>Case C</i> $\sigma_e^2 = 0.01$ (3)	<i>Case D</i> $\sigma_e^2 = 0.005$ (4)
β_0	1.42* (0.19)	1.62* (0.14)	1.25* (0.14)	1.28* (0.11)
β_1	0.46* (0.024)	0.43* (0.02)	0.48* (0.017)	0.46* (0.01)
β_2	0.48* (0.017)	0.47* (0.013)	0.48* (0.01)	0.47* (0.01)
σ_τ	0.15* (0.035)	0.11* (0.01)	0.15* (0.01)	0.15* (0.01)
σ_ν	0.15* (0.01)	0.13* (0.02)	0.08* (0.01)	0.04 (0.025)

The asterisk “*” denotes statistical significance at 0.05 significance level or better. Standard errors are shown in parentheses.

The values for σ_e^2 shown in the top row of the table represent the true variances for the statistical error distributions.

where D represents a dummy variable which is assigned the following values

$$D = \begin{cases} 1 & : \text{if a DMU is identified as 100\% efficient in stage 1} \\ 0 & : \text{if a DMU is not identified as 100\% efficient in stage 1.} \end{cases} \quad (9.21)$$

Tables 9.4 and 9.5 exhibit the results secured when (9.19) was used as the estimating relation to obtain parameter values for both OLS and SF regressions. As might be expected, all parameter estimates are wide of the true values represented in (9.17) and significantly so in all cases as recorded in Tables 9.4 and 9.5. Somewhat surprising, however, is the fact that the OLS and SF estimates are very close and, in many cases they are identical.

When (9.20) is used — which is the regression with dummy variable values described in (9.21) — the situation is reversed for the efficient, but not for the inefficient DMUs. When the estimates are formed in the manner noted at the bottoms of Tables 9.6 and 9.7, none of the estimate of β_1 and β_2 differ significantly from their true values as given in (9.17). These are the estimates to be employed for $D = 1$. For $D = 0$, the case of inefficient DMUs, the previous result is repeated. All of the estimates differ significantly from their true values in both the empirical and simulation studies we described as can be seen in both of Tables 9.6 and 9.7.

Table 9.6. OLS Regression Estimates without Dummy Variables on DEA-efficient DMUsCase 1: *EXPONENTIAL* distribution of input inefficiencies

Parameter Estimates	<i>Case A</i> $\sigma_\epsilon^2 = 0.04$	<i>Case B</i> $\sigma_\epsilon^2 = 0.0225$	<i>Case C</i> $\sigma_\epsilon^2 = 0.01$	<i>Case D</i> $\sigma_\epsilon^2 = 0.005$
	(1)	(2)	(3)	(4)
β_0	1.07* (0.21)	1.47* (0.17)	1.28* (0.14)	1.34* (0.11)
β_1	0.49* (0.03)	0.43* (0.02)	0.46* (0.02)	0.47* (0.01)
β_2	0.48* (0.02)	0.48* (0.015)	0.48* (0.01)	0.46* (0.01)
δ	-1.57* (0.64)	-2.30* (0.43)	-1.50* (0.35)	-1.50* (0.21)
δ_1	0.155* (0.075)	0.26* (0.05)	0.16* (0.04)	0.16* (0.03)
δ_2	0.12* (0.05)	0.12* (0.04)	0.10* (0.03)	0.09* (0.02)

Combining Parameters with <i>Dummy Variables</i>	$H_0 : \beta_1 + \delta_1 = 0.65$	$t_1 = 0.07$	$t_1 = 0.87$	$t_1 = -0.72$	$t_1 = 0.82$
	$H_a : \beta_1 + \delta_1 \neq 0.65$				
	$H_0 : \beta_2 + \delta_2 = 0.55$	$t_2 = 1.09$	$t_2 = 1.76$	$t_2 = 1.02$	$t_2 \simeq 0$
	$H_a : \beta_2 + \delta_2 \neq 0.55$				

The asterisk “*” denotes statistical significance at 0.05 significance level or better. Standard errors are shown in parentheses.

The values for σ_ϵ^2 shown in the top row of the table represent the true variances for the statistical error distributions.

The above tables report results for an exponential distribution of the inefficiencies associated with \hat{x}_1, \hat{x}_2 as defined in (9.18). However, uses of other statistical distributions and other forms of production functions did not alter these kinds of results for either the efficient or the inefficient DMUs. Thus this two-stage approach provided a new way of evaluating efficient and inefficient behaviors in both the empirical and simulation studies where it was used. It also provides an OLS regression as an alternative to the SF regression and this alternative is easier to use (or at least is more familiar) for many uses. See Brockett *et al.* (2004)³⁶ for an application to advertising strategy and a comparison with other types of statistical regression.

Comment : There are shortcomings and research challenges that remain to be met. One such challenge is to expand these uses to include multiple outputs as well as multiple inputs. Another challenge is to develop ways for identifying and estimating input specific as well as output specific inefficiencies. In order to meet such challenges it will be necessary to develop an analytically based theory in order to extend what can be accomplished by empirical applications and simulation studies.

Table 9.7. Stochastic Frontier Regression Estimates without Dummy Variables on DEA-efficient DMUs

Case 1: EXPONENTIAL distribution of input inefficiencies

Parameter Estimates	<i>Case A</i> $\sigma_\epsilon^2 = 0.04$	<i>Case B</i> $\sigma_\epsilon^2 = 0.0225$	<i>Case C</i> $\sigma_\epsilon^2 = 0.01$	<i>Case D</i> $\sigma_\epsilon^2 = 0.005$
	(1)	(2)	(3)	(4)
β_0	1.18* (0.23)	1.50* (0.16)	0.80* (0.16)	1.40* (0.13)
β_1	0.50* (0.03)	0.44* (0.02)	0.53* (0.02)	0.49* (0.01)
β_2	0.48* (0.02)	0.49* (0.02)	0.50* (0.01)	0.47* (0.02)
δ	-1.60* (0.57)	-2.4* (0.56)	-1.25* (0.38)	-1.55* (0.23)
δ_1	0.16* (0.07)	0.26* (0.06)	0.13* (0.04)	0.15* (0.03)
δ_2	0.11* (0.05)	0.13* (0.04)	0.086* (0.04)	0.09* (0.03)
σ_ν	0.13(0.01)*	0.09(0.01)*	0.05(0.01)*	0.04(0.01)*
<hr/>				
Combining Parameters with <i>Dummy Variables</i>				
$H_0 : \beta_1 + \delta_1 = 0.65$ $t_1 = 0.20$ $t_1 = 0.93$ $t_1 = 0.28$ $t_1 = -0.4$				
$H_a : \beta_1 + \delta_1 \neq 0.65$				
$H_0 : \beta_2 + \delta_2 = 0.55$ $t_2 = 1.03$ $t_2 = 1.90$ $t_2 = 1.16$ $t_2 = 0.45$				
$H_a : \beta_2 + \delta_2 \neq 0.55$				

The asterisk “*” denotes statistical significance at 0.05 significance level or better. Standard errors are shown in parentheses.

The values for σ_ϵ^2 shown in the top row of the table represent the true variances for the statistical error distributions.

Fortunately, excellent texts dealing with stochastic frontier and other approaches to efficiency evaluation have become available in the following two books,

1. T. Coelli, D.S.P. Rao and G.E. Battese (1998) *An Introduction to Efficiency and Productivity Analysis* (Boston: Kluwer Academic Publishers).
2. S.C. Kumbhakar and C.A.K. Lovell (2000) *Stochastic Frontier Analysis* (Cambridge: Cambridge University Press).

9.4 CHANCE-CONSTRAINED PROGRAMMING AND SATISFICING IN DEA

9.4.1 Introduction

S. Thore's (1987)³⁷ paper initiated a series of joint efforts with R. Land, and C.A.K. Lovell³⁸ directed to joining chance-constrained programming (CCP) with DEA as a third method for treating data uncertainties in DEA. Here we turn to Cooper, Huang and Li (1996)³⁹ to show how this approach can also be used to make contact with the concept of "satisficing" as developed in the psychology literature by H.A. Simon⁴⁰ as an alternative to the assumption of "optimizing" behavior which is extensively used in economics.

9.4.2 Satisficing in DEA

We start with the following CCP formulation that extends the CCR (ratio) model of DEA which was introduced in Section 2.3 of Chapter 2,

$$\begin{aligned} \max \quad & P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{ro}}{\sum_{i=1}^m v_i \tilde{x}_{io}} \geq \beta_o\right) \\ \text{subject to} \quad & P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{rj}}{\sum_{i=1}^m v_i \tilde{x}_{ij}} \leq \beta_j\right) \geq 1 - \alpha_j, \quad j = 1, \dots, n \\ & u_r, v_i \geq 0 \quad \forall r, i. \end{aligned} \tag{9.22}$$

Here " P " means "probability" and " \sim " identifies these outputs and inputs as random variables with a known probability distribution while $0 \leq \alpha_j \leq 1$ is a scalar, specified in advance, which represents an allowable chance (=risk) of failing to satisfy the constraints with which it is associated.

For "satisficing," the values of β_o is interpreted as an "aspiration level" specified as an efficiency rating which is to be attained. The β_j are also prescribed constants imposed by the individual, or by outside conditions including superior levels of management.

To exhibit another aspect of satisficing behavior we might consider the case of inconsistent constraints. The problem will then have no solution. In such cases, according to Simon, an individual must either quit or else he must revise his aspiration level — or the risk of not achieving this level (or both). Thus, probabilistic (chance-constrained programming) formulations allow for types

of behavior which are not present in the customary deterministic models of satisfying.

Now, however, we want to make contact with the deterministic DEA models which we discussed earlier. For this purpose we select the CCR model which was introduced as early as Chapter 2. This model always has a solution and the same is true for (9.22). This can be exhibited by choosing $u_r = 0 \forall r$ and $v_i > 0$ for some i . Although not minimal for the objective, this choice satisfies all constraints with a probability of unity.

9.4.3 Deterministic Equivalents

To align the development more closely with our earlier versions of the CCR ratio model we note that

$$P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{ro}}{\sum_{i=1}^m v_i \tilde{x}_{io}} \leq \beta_o\right) + P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{ro}}{\sum_{i=1}^m v_i \tilde{x}_{io}} \geq \beta_o\right) = 1 \quad (9.23)$$

where, for simplicity, we restrict attention to the class of continuous distributions. We therefore replace (9.22) with

$$\begin{aligned} & \max \quad P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{ro}}{\sum_{i=1}^m v_i x_{io}} \geq \beta_o\right) \\ \text{subject to} \quad & P\left(\frac{\sum_{r=1}^s u_r \tilde{y}_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq \beta_j\right) \geq 1 - \alpha_j, \quad j = 1, \dots, n \\ & u_r, v_i \geq 0 \quad \forall r, i. \end{aligned} \quad (9.24)$$

Here we follow Land, Lovell and Thore (1993) and omit the symbol “~” from the x_{ij} (and x_{io}) in order to represent the inputs as deterministic. This model corresponds to a situation in which DMU managers choose the inputs without being able to completely control the outputs. Moreover if we also remove the symbol “~” from the y_{rj} (and y_{ro}), set $\beta_j = 1$, $j = 1, \dots, n$ and remove β_o from the objective we will reproduce the CCR model that was first encountered in expression (2.3)-(2.6) in Chapter 2.

This identification having been made, we restore the symbol “~” to the y_{rj} (and y_{ro}) — thereby characterizing them as random variables — but we continue to treat the x_{ij} (and x_{io}) as deterministic. Then using vector-matrix notation we restate the constraints in (9.24) via the following development,

$$P\left(\frac{\mathbf{u}^T \tilde{\mathbf{y}}_j}{\mathbf{v}^T \mathbf{x}_j} \leq \beta_j\right) = P(\mathbf{u}^T \tilde{\mathbf{y}}_j \leq \beta_j \mathbf{v}^T \mathbf{x}_j). \quad (9.25)$$

Now let $\bar{\mathbf{y}}_j$ be the vector of output means and let Σ_j represent the variance-covariance matrix. We assume that this matrix is positive definite so we can represent the variance by $\mathbf{u}^T \Sigma_j \mathbf{u}$, a scalar, which is also positive for all choices of $\mathbf{u} \neq \mathbf{0}$. We then subtract $\mathbf{u}^T \bar{\mathbf{y}}_j$ from both sides of the right-hand inequality

in (9.25) and divide through by $\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}$ to obtain

$$P \left(\frac{\mathbf{u}^T \tilde{\mathbf{y}}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}} \leq \frac{\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}} \right) \geq 1 - \alpha_j \quad (9.26)$$

for each $j = 1, \dots, n$. Now we note that the expression on the right in the parenthesis does not contain any random elements.

To simplify our notation we introduce a new random variable defined by

$$\tilde{z}_j = \frac{\mathbf{u}^T \tilde{\mathbf{y}}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}}. \quad (9.27)$$

We then replace (9.26) with

$$P(\tilde{z}_j \leq k_j(\mathbf{u}^T, \mathbf{v}^T)) \geq 1 - \alpha_j, \quad j = 1, \dots, n \quad (9.28)$$

where

$$k_j(\mathbf{u}^T, \mathbf{v}^T) = \frac{\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}}$$

so we can write

$$\int_{-\infty}^{k_j(\mathbf{u}^T, \mathbf{v}^T)} f(z_j) dz_j = \Phi \left(\frac{\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}} \right) \geq 1 - \alpha_j \quad (9.29)$$

in place of (9.26).

We now assume that Φ is the normal distribution which has been standardized via

$$\tilde{z}_j = \frac{\mathbf{u}^T (\tilde{\mathbf{y}}_j - \bar{\mathbf{y}}_j)}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}}. \quad (9.30)$$

Assuming $\alpha_j \leq 0.5$ we can utilize the property of invertibility associated with this distribution and apply it to (9.29) to obtain

$$\frac{\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j}{\sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}} \geq \Phi^{-1}(1 - \alpha_j) \quad (9.31)$$

where Φ^{-1} is the fractile function associated with the standard normal distribution. Hence also

$$\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j \geq \Phi^{-1}(1 - \alpha_j) \sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}. \quad (9.32)$$

We now employ what Charnes and Cooper (1963)⁴¹ refer to as “splitting variables” which we symbolize by η_j in

$$\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j \geq \eta_j \geq \Phi^{-1}(1 - \alpha_j) \sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}}. \quad (9.33)$$

For every $j = 1, \dots, n$ this variable is nonnegative by virtue of the expression on the right. Provided this nonnegativity is preserved we can therefore use this variable to split the expression in (9.33) into the following pair

$$\begin{aligned}\beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j &\geq \eta_j \geq 0 \\ K_{(1-\alpha_j)}^2 \mathbf{u}^T \Sigma_j \mathbf{u} &\leq \eta_j^2\end{aligned}\quad (9.34)$$

where

$$\begin{aligned}K_{(1-\alpha_j)} &= \Phi^{-1}(1 - \alpha_j) \\ j &= 1, \dots, n.\end{aligned}$$

We have thus separated the conditions in (9.33) into a pair for which the first relation refers to a valuation effected by multipliers assigned to the inputs and outputs while the second relation treats the “risks” as in a portfolio analysis of the Markowitz-Sharpe type used in finance.⁴²

In place (9.24) we now have

$$\begin{aligned}\max \quad & P\left(\frac{\mathbf{u}^T \tilde{\mathbf{y}}_o}{\mathbf{v}^T \mathbf{x}_o} \geq \beta_o\right) \\ \text{subject to} \quad & \beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j - \eta_j \geq 0 \\ & K_{(1-\alpha_j)}^2 \mathbf{u}^T \Sigma_j \mathbf{u} - \eta_j^2 \leq 0 \\ & \mathbf{v}, \mathbf{u} \geq \mathbf{0}, \eta_j \geq 0, j = 1, \dots, n.\end{aligned}\quad (9.35)$$

The constraints, but not the objective, are now deterministic. To bring our preceding development to bear we therefore replace (9.35) with

$$\begin{aligned}\max \quad & \gamma_o \\ \text{subject to} \quad & P\left(\frac{\mathbf{u}^T \tilde{\mathbf{y}}_o}{\mathbf{v}^T \mathbf{x}_o} \geq \beta_o\right) \geq \gamma_o \\ & \beta_j \mathbf{v}^T \mathbf{x}_j - \mathbf{u}^T \bar{\mathbf{y}}_j - \eta_j \geq 0 \\ & K_{(1-\alpha_j)}^2 \mathbf{u}^T \Sigma_j \mathbf{u} - \eta_j^2 \leq 0 \\ & \mathbf{v}, \mathbf{u} \geq \mathbf{0}, \eta_j \geq 0, j = 1, \dots, n.\end{aligned}\quad (9.36)$$

Proceeding as before we then have

$$\begin{aligned}\max \quad & \gamma_o \\ \text{subject to} \quad & \mathbf{u}^T \bar{\mathbf{y}}_o - \beta_o \mathbf{v}^T \mathbf{x}_o \geq \Phi^{-1}(\gamma_o) \sqrt{\mathbf{u}^T \Sigma_j \mathbf{u}} \\ & \eta_j + \mathbf{u}^T \mathbf{y}_j - \beta_j \mathbf{v}^T \mathbf{x}_j \leq 0 \\ & \eta_j^2 - K_{(1-\alpha_j)}^2 \mathbf{u}^T \Sigma_j \mathbf{u} \geq 0 \\ & \mathbf{u}, \mathbf{v} \geq \mathbf{0}, \eta_j \geq 0, j = 1, \dots, n \\ & 0 \leq \gamma_o \leq 1.\end{aligned}\quad (9.37)$$

This is a “deterministic equivalent” for (9.24) in that the optimal values of \mathbf{u}^* , \mathbf{v}^* in (9.37) will also be optimal for (9.24).

9.4.4 Stochastic Efficiency

Although entirely deterministic, this problem is difficult to solve because, by virtue of the first constraint, it is nonconvex as well as nonlinear. As shown in Cooper, Huang and Li (1996), it is possible to replace (9.37) with a convex programming problem but we do not undertake the further development needed to show this. Instead, we assume that we have a solution with

$$\gamma_o^* = P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \geq \beta_o \right), \quad (9.38)$$

where γ_o^* is obtained from (9.36).

To develop the concepts of stochastic efficiency we assume that $\beta_o = \beta_{j_o}$ so the level prescribed for DMU_o in its constraints is the same as the β_o level prescribed in the objective. Then we note that $\gamma_o^* > \alpha_{j_o}$ is not possible because this would fail to satisfy this constraint. To see that this is so we note, as in (9.23), that

$$P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \geq \beta_o \right) + P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \leq \beta_o \right) = 1 \quad (9.39)$$

because $P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} = \beta_o \right) = 0$ for a continuous distribution. Hence

$$\begin{aligned} P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \leq \beta_o \right) &= 1 - P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \geq \beta_o \right) \\ &= 1 - \gamma_o^* < 1 - \alpha_{j_o} \end{aligned} \quad (9.40)$$

which fails to satisfy the constraint

$$P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \leq \beta_o \right) \geq 1 - \alpha_{j_o}.$$

Now if $\beta_{j_o} = \beta_o = 1$, which will usually be the case of interest, the above development leads to the following,

Theorem 9.1 (Cooper, Huang and Li (1996)) *If $\beta_{j_o} = \beta_o = 1$ then DMU_o will have performed in a stochastically efficient manner if and only if $\gamma_o^* = \alpha_{j_o}$.*

This leaves the case of $\gamma_o^* < \alpha_{j_o}$ to be attended to. In this case the risk of failing to satisfy the constraints for DMU_{j_o} falls below the level which was specified as satisfactory. To restate this in a more positive manner, we return to (9.40) and reorient it to

$$\begin{aligned} P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \leq \beta_o \right) &= 1 - P \left(\frac{\mathbf{u}^{*T} \tilde{\mathbf{y}}_o}{\mathbf{v}^{*T} \mathbf{x}_o} \geq \beta_o \right) \\ &= 1 - \gamma_o^* > 1 - \alpha_{j_o}. \end{aligned} \quad (9.41)$$

This leads to the following corollary to the above theorem,

Corollary 9.1 If $\beta_{j_o} = \beta_o = 1$ then DMU_o 's performance is stochastically inefficient with probability $1 - \gamma_o^*$ if and only if $\gamma_o^* < \alpha_{j_o}$.

We justify this by noting that (9.41) means that the probability of falling below β_o exceeds the probability that was specified as being satisfactory.

To return to the satisficing model and concepts discussed in Section 9.4.2, we assume that $0 \leq \beta_o = \beta_{j_o} < 1$. The above theorem and corollary then translate into: "satisficing was attained or not according to whether $\gamma_o^* = \alpha_{j_o}$ or $\gamma_o^* < \alpha_{j_o}$." To see what this means we note that (9.39) consists of opposed probability statements except for the case $P\left(\frac{\mathbf{u}^{*T}\tilde{\mathbf{y}}_o}{\mathbf{v}^{*T}\tilde{\mathbf{x}}_o} = \beta_o\right) = 0$. Hence failure to attain satisficing occurs when the goal specified in the objective can be attained only at a level below the risk that is specified for being wrong in making this inference.

Returning to our earlier choice of $u_r = 0 \forall r$ and some $v_i > 0$ for (9.22) we note that this assigns a positive value to some inputs and a zero value for all outputs. The objective and the associated constraints in (9.22) will then be achieved with probability one because of refusal to play and hence there will be a zero probability of achieving the objective in (9.24). See (9.39). This is an important special case of Simon's "refusal to play" behavior that was noted in the discussion following (9.22).

Comment : This is as far as we carry our analyses of these chance constrained programming approaches to DEA. We need to note, however, that this analysis has been restricted to what is referred to as the "P-model" in the chance-constrained programming literature. Most of the other DEA literature on this topic has utilized the "E-model," so named because its objective is stated in terms of optimizing "expected values." None has yet essayed a choice of "V-models" for which the objective is to minimize "variance" or "mean-square error." See A. Charnes and W.W. Cooper (1963) "Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints," *Operations Research* 11, pp.18-39.

The formulations here (and elsewhere in the DEA literature) are confined to a use of "conditional chance constraints." A new chapter was opened for this research, however, by Olesen and Petersen (1995)⁴³ who used "joint chance constraints" in their CCP models. In a subsequent paper, Cooper, Huang, Lelas, Li and Olesen (1998)⁴⁴ utilize such joint constraints to extend the concept of "stochastic efficiency" to a measure called " α -stochastic efficiency" for use in problems where "efficiency dominance" is of interest.

Like the rest of the DEA literature dealing with CCP, we have restricted attention to the class of "zero order decision rules." This corresponds to a "here and now" approach to decision making in contrast to the "wait and see" approach that is more appropriate to dynamic settings in which it may be better to delay some parts of a decision until more information is available. To go further in this direction leads to the difficult problem of choosing a best decision rule from an admissible class of decision rules. To date, this problem has only been addressed in any detail for the class of linear (first order) decision rules

even in the general literature on CCP and even this treatment was conducted under restrictive assumptions on the nature of the random variables and their statistical distributions.⁴⁵ See also Charnes, Cooper and Symods (1958),⁴⁶ an article which originated the (as-yet-to-be named) "chance constraint programming." This topic is important and invites further treatments as an example of the "information processing" that plays an important role in studying use of the term "satisficing behavior." See Gigerenzer (2004).⁴⁷ Finally we come to the treatments which, by and large, have assumed that these probability distributions are known. Hence there is a real need and interest in relaxing this assumption. Only bare beginnings on this topic have been made as in R. Jagannathan (1985) "Use of Sample Information in Stochastic Recourse and Chance Constrained Programming Models," *Management Science* 31, pp.96-108.

9.5 SUMMARY OF CHAPTER 9

In this chapter we have treated the topic of data variability in the following manner. Starting with the topic of sensitivity and stability analysis we moved to statistical regression approaches which we aligned with DEA in various ways. We then went on to probabilistic formulations using the P-model of chance-constrained programming which we could relate to our CCR and BCC models which we had previously treated in deterministic manners.

The topics treated in this chapter are all under continuing development. Hence we do not provide problems and suggested answers like those we presented in preceding chapters. Instead we have provided comments and references that could help to point up issues for further research. We hope that readers will respond positively and regard this as an opportunity to join in the very active research that is going on in these (and other) areas.

We here note that research on the topics treated in this chapter were prompted by problems encountered in attempts to bring a successful conclusion to one or many attempts to use DEA in actual applications. Indeed a good deal of the very considerable progress in DEA has emanated from actual attempts to apply it to different problems. This, too, has been an important source of progress in that new applications as well as new developments in DEA are being simultaneously reported.

To help persons who want to pursue additional topics and uses we have supplied an extensive bibliography in the disk that accompanies this book as well as in this chapter. We hope this will be helpful and we hope readers of our text will experience some of the fun and exhilaration that we have experienced as we watch the rapid pace of developments in DEA.

Notes

1. P.W. Wilson (1995), "Detecting Influential Observations in Data Envelopment Analysis," *Journal of Productivity Analysis* 6, pp.27-46.
2. See, for instance, R.M. Thrall (1989), "Classification of Transitions under Expansion of Inputs and Outputs," *Managerial and Decision Economics* 10, pp.159-162.

3. R.D. Banker, H. Chang and W.W. Cooper (1996), "Simulation Studies of Efficiency, Returns to Scale and Misspecification with Nonlinear Functions in DEA," *Annals of Operations Research* 66, pp.233-253.
4. A. Charnes, W.W. Cooper, A.Y. Lewin, R.C. Morey and J.J. Rousseau (1985), "Sensitivity and Stability Analysis in DEA," *Annals of Operations Research* 2, pp.139-156.
5. A. Charnes and W.W. Cooper (1968), "Structural Sensitivity Analysis in Linear Programming and an Exact Product Form Left Inverse," *Naval Research Logistics Quarterly* 15, pp.517-522.
6. For a summary discussion see A. Charnes and L. Neralic (1992), "Sensitivity Analysis in Data Envelopment Analysis 3," *Glasnik Matematicki* 27, pp.191-201. A subsequent extension is L. Neralic (1997), "Sensitivity in Data Envelopment Analysis for Arbitrary Perturbations of Data," *Glasnik Matematicki* 32, pp.315-335. See also L. Neralic (2004), "Preservation of Efficiency and Inefficiency Classification in Data Envelopment Analysis," *Mathematical Communications* 9, pp.51-62.
7. A. Charnes, S. Haag, P. Jaska and J. Semple (1992), "Sensitivity of Efficiency Calculations in the Additive Model of Data Envelopment Analysis," *International Journal of System Sciences* 23, pp.789-798. Extensions to other classes of models may be found in A. Charnes, J.J. Rousseau and J.H.Semple (1996) "Sensitivity and Stability of Efficiency Classifications in DEA," *Journal of Productivity Analysis* 7, pp.5-18.
8. The shape of this "ball" will depend on the norm that is used. For a discussion of these and other metric concepts and their associated geometric portrayals see Appendix A in A. Charnes and W.W. Cooper (1961), *Management Models and Industrial Applications of Linear Programming* (New York: John Wiley & Sons).
9. This omission of DMU_0 is also used in developing a measure of "super efficiency" as it is called in P. Andersen and N.C. Petersen (1993), "A Procedure for Ranking Efficient Units in DEA," *Management Science* 39, pp.1261-1264. Their use is more closely associated with stability when a DMU is omitted, however, so we do not cover it here. See the next chapter, Chapter 10, in this text. See also R.M. Thrall (1996) "Duality Classification and Slacks in DEA," *Annals of Operations Research* 66, pp.104-138.
10. A. Charnes, J.J. Rousseau and J.H. Semple (1996), "Sensitivity and Stability of Efficiency Classification in Data Envelopment Analysis," *Journal of Productivity Analysis* 7, pp.5-18.
11. In fact, Seiford and Zhu propose an iterative approach to assemble an exact stability region in L. Seiford and J. Zhu, "Stability Regions for Maintaining Efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 108, 1998, pp.127-139.
12. R.G. Thompson, P.S. Dharmpala and R.M. Thrall (1994), "Sensitivity Analysis of Efficiency Measures with Applications to Kansas Farming and Illinois Coal Mining," in A. Charnes, W.W. Cooper, A.Y. Lewin and L.M. Seiford, eds., *Data Envelopment Analysis: Theory, Methodology and Applications* (Norwell, Mass., Kluwer Academic Publishers) pp.393-422.
13. R.G. Thompson, P.S. Dharmpala, J. Diaz, M.D. Gonzales-Lina and R.M. Thrall (1996), "DEA Multiplier Analytic Center Sensitivity Analysis with an Illustrative Application to Independent Oil Cos.," *Annals of Operations Research* 66, pp.163-180.
14. A. Charnes, W.W. Cooper and R.M.Thrall (1991), "A Structure for Classifying and Characterizing Efficiency in Data Envelopment Analysis," *Journal of Productivity Analysis* 2, pp.197-237.
15. An alternate approach to simultaneous variations in all data effected by the envelopment model is available in L.M. Seiford and J. Zhu (1998), "Sensitivity Analysis of DEA Models for Simultaneous Changes in All Data," *Journal of the Operational Research Society* 49, pp.1060-1071. See also the treatments of simultaneous changes in all data for additive models in L. Neralic (2004), "Preservation of Efficiency and Inefficiency Classification in Data Envelopment Analysis," *Mathematical Communications* 9, pp.51-62.
16. The ranges within which these dual variable values do not change generally form part of the printouts in standard linear programming computer codes.

17. This status is easily recognized because $\theta_1^* = \theta_2^* = \theta_3^* = 1$ are all associated with uniquely obtained solutions with zero slacks. See A. Charnes, W.W. Cooper and R.M. Thrall (1991) "A Structure for Classifying and Characterizing Efficiency in Data Envelopment Analysis," *Journal of Productivity Analysis* 2, pp.197-237.
18. W.W. Cooper, S. Li, L.M. Seiford and J. Zhu (2004), Chapter 3 in W.W. Cooper, L.M. Seiford and J. Zhu, eds., *Handbook on Data Envelopment Analysis* (Norwell, Mass., Kluwer Academic Publishers).
19. R.D. Banker (1993), "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation," *Management Science* 39, pp.1265-1273.
20. R. Banker and R. Natarajan (2004), "Statistical Tests Based on DEA Efficiency Scores," Chapter 11 in in W.W. Cooper, L.M. Seiford and J. Zhu, eds., *Handbook on Data Envelopment Analysis* (Norwell, Mass., Kluwer Academic Publishers).
21. Quoted from p.139 in R.D Banker (1996), "Hypothesis Tests Using Data Envelopment Analysis," *Journal of Productivity Analysis* pp.139-159.
22. A.P. Korostolev, L. Simar and A.B. Tsybakov (1995), "On Estimation of Monotone and Convex Boundaries," *Public Institute of Statistics of the University of Paris*, pp.3-15. See also Korostolev, Simar and Tsybakov (1995), "Efficient Estimation of Monotone Boundaries," *Annals of Statistics* 23, pp.476-489.
23. See the discussion in L. Simar (1996), "Aspects of Statistical Analysis in DEA-Type Frontier Models," *Journal of Productivity Analysis* 7, pp.177-186.
24. L. Simar and P.W. Wilson (1998), "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models," *Management Science* 44, pp.49-61. See also Simar and Wilson (2004) "Performance of the Bootstrap for DEA Estimators and Iterating the Principle," Chapter 10 in W.W. Cooper, L.M. Seiford and J. Zhu, eds. *Handbook on Data Envelopment Analysis* (Norwell, Mass: Kluwer Academic Publishers).
25. M.J. Farrell (1951), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society Series A*, 120, pp.253-290.
26. D.J. Aigner and S.F. Chu (1968), "On Estimating the Industry Production Frontiers," *American Economic Review* 56, pp.826-839.
27. D.J. Aigner, C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Models," *Journal of Econometrics* 6, pp.21-37. See also W. Meeusen and J. van den Broeck (1977) "Efficiency Estimation from Cobb-Douglas Functions with Composed Error," *International Economic Review* 18, pp.435-444.
28. J. Jondrow, C.A.K. Lovell, I.S. Materov and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model," *Journal of Econometrics* 51, pp.259-284.
29. B.H. Gong and R.C. Sickles (1990), "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data," *Journal of Econometrics* 51, pp.259-284.
30. P. Schmidt (1985-1986), "Frontier Production Functions," *Econometric Reviews* 4, pp.289-328. See also P.W. Bauer (1990), "Recent Development in Econometric Estimation of Frontiers," *Journal of Econometrics* 46, pp.39-56.
31. G.D. Ferrier and C.A.K. Lovell (1990), "Measuring Cost Efficiency in Banking — Econometric and Linear Programming Evidence," *Journal of Econometrics* 6, pp.229-245.
32. A. Charnes, W.W. Cooper and T. Sueyoshi (1988), "A Goal Programming/Constrained Regression Review of the Bell System Breakup," *Management Science* 34, pp.1-26.
33. See R.S. Barr, L.M. Seiford and T.F. Siems (1994), "Forecasting Bank Failure: A Non-Parametric Frontier Estimation Approach," *Recherches Economiques de Louvain* 60, pp. 417-429. for an example of a different two-stage DEA regression approach in which the DEA scores from the first stage served as an independent variable in the second stage regression model.
34. V. Arnold, I.R. Bardhan, W.W. Cooper and S.C. Kumbhakar (1994), "New Uses of DEA and Statistical Regressions for Efficiency Evaluation and Estimation — With an Illustrative Application to Public Secondary Schools in Texas," *Annals of Operations Research* 66, pp.255-278.

35. I.R. Bardhan, W.W. Cooper and S.C. Kumbhakar (1998), "A Simulation Study of Joint Uses of Data Envelopment Analysis and Stochastic Regressions for Production Function Estimation and Efficiency Evaluation," *Journal of Productivity Analysis* 9, pp.249-278.
36. P.L. Brockett, W.W. Cooper, S.C. Kumbhakar, M.J. Kwinn Jr. and D. McCarthy (2004), "Alternative Statistical Regression Studies of the Effects of Joint and Service-Specific Advertising on Military Recruitment," *Journal of the Operational Research Society* 55, pp.1039-1048.
37. S. Thore (1987), "Chance-Constrained Activity Analysis," *European Journal of Operational Research* 30, pp.267-269.
38. See the following three papers by K.C. Land, C.A.K. Lovell and S. Thore: (1) "Productive Efficiency under Capitalism and State Socialism: the Chance Constrained Programming Approach" in Pierre Pestieau, ed. in *Public Finance in a World of Transition* (1992) supplement to *Public Finance* 47, pp.109-121; (2) "Chance-Constrained Data Envelopment Analysis," *Managerial and Decision Economics* 14, 1993, pp.541-554; (3) "Productive Efficiency under Capitalism and State Socialism: An Empirical Inquiry Using Chance-Constrained Data Envelopment Analysis," *Technological Forecasting and Social Change* 46, 1994, pp.139-152. In "Four Papers on Capitalism and State Socialism" (Austin Texas: The University of Texas, IC² Institute) S. Thore notes that publication of (2) was delayed because it was to be presented at a 1991 conference in Leningrad which was cancelled because of the Soviet Crisis.
39. W.W. Cooper, Z. Huang and S. Li (1996), "Satisficing DEA Models under Chance Constraints," *Annals of Operations Research* 66, pp.279-295. For a survey of chance constraint programming uses in DEA, see W.W. Cooper, Z. Huang and S. Li (2004), "Chance Constraint DEA," in W.W. Cooper, R.M. Seiford and J. Zhu, eds., *Handbook on Data Envelopment Analysis* (Norwell, Mass., Kluwer Academic Publishers).
40. See Chapter 15 in H.A. Simon (1957), *Models of Man* (New York: John Wiley & Sons, Inc.)
41. A. Charnes and W.W. Cooper (1963), "Deterministic Equivalents for Optimizing and Satisficing under Chance Constraints," *Operations Research* 11, pp.18-39.
42. See W.F. Sharpe (1970), *Portfolio Theory and Capital Markets* (New York: McGraw Hill, Inc.)
43. O.B. Olesen and N.C. Petersen (1995), "Chance Constrained Efficiency Evaluation," *Management Science* 41, pp.442-457.
44. W.W. Cooper, Z. Huang, S.X. Li and O.B. Olesen (1998), "Chance Constrained Programming Formulations for Stochastic Characterizations of Efficiency and Dominance in DEA," *Journal of Productivity Analysis* 9, pp.53-79.
45. See A. Charnes and W.W. Cooper (1963) in footnote 41, above.
46. A. Charnes, W.W. Cooper and G.H. Symods (1958), "Cost Horizons and Certainty Equivalents," *Management Science* 4, pp.235-263.
47. G. Gigerenzer (2004), "Striking a Blow for Sanity in Theories of Rationality," in M. Augier and J.G. March, eds., *Models of a Man: Essays in Memory of H.A. Simon* (Cambridge: MIT Press).

10 SUPER-EFFICIENCY MODELS

10.1 INTRODUCTION

In this chapter we introduce a model proposed by Andersen and Petersen (1993),¹ that leads to a concept called “super-efficiency.” The efficiency scores from these models are obtained by eliminating the data on the DMU_o to be evaluated from the solution set. For the input model this can result in values which are regarded as according DMU_o the status of being “super-efficient.” These values are then used to rank the DMUs and thereby eliminate some (but not all) of the ties that occur for efficient DMUs.

Other uses of this approach have also been proposed. Wilson (1993)², for example, suggests two uses of these measures in which each DMU_o is ranked according to its “influence” in either (or both) of the following two senses: (1) the number of observations that experience a change in their measure of technical efficiency as a result of these eliminations from the solution set and (2) the magnitude of these changes. Still other interpretations and uses are possible, and we shall add further to such possibilities. See also Ray (2000).³

There are troubles with these “super-efficiency” measures, as we shall see. These troubles can range from a lack of units invariance for these measures and extend to non-solution possibilities when convexity constraints are to be dealt with — as in the BCC model. However, the underlying concept is important, so we shall first review this approach to ranking in the form suggested by Andersen

and Petersen (1993) and then show how these deficiencies may be eliminated by using other non-radial models that we will suggest.

10.2 RADIAL SUPER-EFFICIENCY MODELS

We start with the model used by Andersen and Petersen which takes the form of a CCR model and thereby avoids the possibility of non-solution that is associated with the convexity constraint in the BCC model. In vector form this model is

$$\begin{aligned}
 [\text{Super Radial-I-C}] \quad \theta^* &= \min_{\theta, \lambda, s^-, s^+} \theta - \varepsilon e s^+ \quad (10.1) \\
 \text{subject to} \quad \theta \mathbf{x}_o &= \sum_{j=1, \neq o}^n \lambda_j \mathbf{x}_j + s^- \\
 \mathbf{y}_o &= \sum_{j=1, \neq o}^n \lambda_j \mathbf{y}_j - s^+
 \end{aligned}$$

where all components of the λ , s^- and s^+ are constrained to be non-negative, $\varepsilon > 0$ is the usual non-Archimedean element and e is a row vector with unity for all elements.

We refer to (10.1) as a “Radial Super-Efficiency” model and note that the vectors, \mathbf{x}_o , \mathbf{y}_o are omitted from the expression on the right in the constraints. The data associated with the DMU_{*o*} being evaluated on the left is therefore omitted from the production possibility set. However, solutions will always exist so long as all elements are positive in the matrices $X, Y > 0$. (For weaker conditions see Charnes, Cooper and Thrall (1991)⁴).

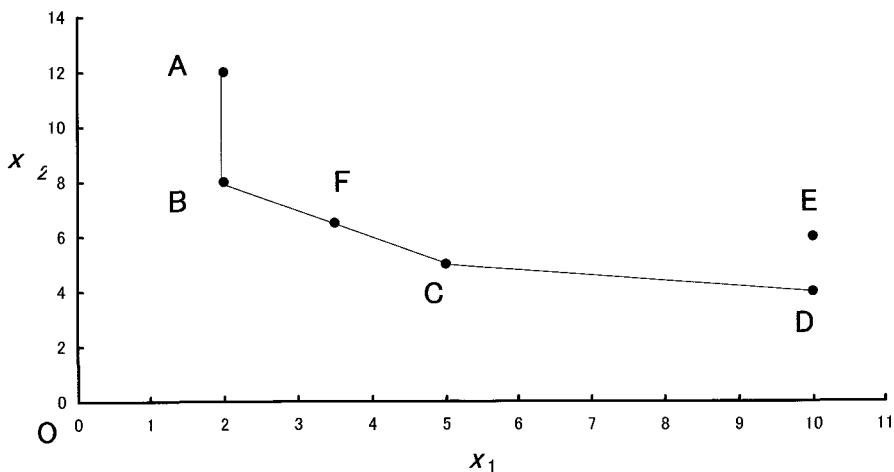
The above model is a member of the class of Input Oriented-CCR (CCR-I) models. The output oriented version (Radial Super-O-C) has an optimal $\phi^* = 1/\theta^*$ and $\lambda^*, s^{-*}, s^{+*}$ adjusted by division with θ^* , so we confine our discussion to the input oriented version, after which we will turn to other models, such as the BCC class of models, where this reciprocal relation for optimal solutions does not hold.

We illustrate the use of (10.1) with the data in Table 10.1 that is represented geometrically in Figure 10.1. Taken from Andersen and Petersen, these data are modified by adding DMU F, (which lies halfway between B and C), to portray all possibilities.

Table 10.1. Test Data

DMU	A	B	C	D	E	F
Input 1	2.0	2.0	5.0	10.0	10.0	3.5
Input 2	12.0	8.0	5.0	4.0	6.0	6.5
Output 1	1.0	1.0	1.0	1.0	1.0	1.0

Source: Andersen and Petersen (1993)

**Figure 10.1.** The Unit Isoquant Spanned by the Test Data in Table 10.1

Consider, first, the evaluation of DMU A as determined from the following adaptation of (10.1) where, as can be seen, the data for A are represented on the left but not on the right.

$$\begin{aligned}
 z^* &= \min \theta - \varepsilon(s_1^- + s_2^-) - \varepsilon s^+ \\
 \text{subject to} \\
 2\theta &= 2\lambda_B + 5\lambda_C + 10\lambda_D + 10\lambda_E + 3.5\lambda_F + s_1^- \\
 12\theta &= 8\lambda_B + 5\lambda_C + 4\lambda_D + 6\lambda_E + 6.5\lambda_F + s_2^- \\
 1 &= \lambda_B + \lambda_C + \lambda_D + \lambda_E + \lambda_F + s^+ \\
 0 \leq \lambda_B, \dots, \lambda_F, s_1^-, s_2^-, s^+ &
 \end{aligned}$$

The solution to this problem is $\theta^* = 1$, $\lambda_B^* = 1$, $s_2^{-*} = 4$. All other variables are zero so $z^* = 1 - 4\varepsilon$. This is the same solution that would be obtained if the data for DMU A were included on the right. The evaluation of an inefficient

point, like A, is not affected by this omission from the production possibility set because the efficient points that enter into the evaluation are unaffected by such a removal.

The latter condition is also present in the case of F. That is, the efficient points used in the evaluation of F are not removed from the production possibility set. Thus, the solution $\theta^* = 1$ and $\lambda_B^* = \lambda_C^* = 1/2$ confirms the efficiency of F both before and after such a removal. The important consideration is that the same efficient points that enter into (or can enter into) the evaluation of a DMU_o are present in both cases. This will also be the case for any inefficient point. For example the evaluation of E is unaffected with $\theta^* = 3/4$ and⁵ $\lambda_C^* = \lambda_D^* = 1/2$ both before and after removal of the point E from the production possibility set. A point like A, which is on a part of the frontier that is not efficient will have the value $\theta^* = 1$, as is also true for the efficient point F which is on a part of the efficient frontier but is not an extreme point. See Charnes, Cooper and Thrall (1991) for a discussion of all possible classes of points and their properties. Now consider the rankings. The slacks are represented in the objective of (10.1) so we have the ranking: DMU F > DMU A since $1 > 1 - 4\epsilon$ while $\theta^* = 3/4$ lies below both values by the definition of $\epsilon > 0$ so this ranking will be F > A > E.

We next delete C from the production possibility set and obtain the solution $\theta^* = 1.2$ and $\lambda_B^* = \lambda_D^* = 1/2$ so that DMU C ranks as a super-efficient convex combination of B and D. This solution gives the coordinates of C with $x_1 = x_2 = 6$ so the elimination of C results in a replacement that yields the same unit of output with a 1 unit (=20%) increase in each input.

The Andersen-Petersen rankings are exhibited in Table 10.2 with B ranked first followed by D. A possible interpretation of what kind of managerial decision might be involved is to assume that the rankings are directed to studying the consequences of eliminating some DMUs. For this purpose, D might be preferred over B because of the large (7.5 units) reduction in input1 — a reduction that is given the negligible weight of $\epsilon > 0$ in the rankings associated with the Andersen-Petersen measure.

Table 10.2. Andersen-Petersen Ranking*

Rank Order	B >	D >	C >	F >	A >	E
z^*	1.32	$1.25 - 7.5\epsilon$	1.2	1	$1 - 4\epsilon$	0.75
Reference Set	$\lambda_A = 0.79$ $\lambda_C = 0.21$	$\lambda_C = 1$	$\lambda_A = 0.5$ $\lambda_D = 0.5$	$\lambda_B = 0.5$ $\lambda_C = 0.5$	$\lambda_B = 1$	$\lambda_C = 0.5$ $\lambda_D = 0.5$
s	0	$s_1^- = 7.5$	0	0	$s_2^- = 4$	0

* Source: Andersen and Petersen (1993)

For purposes like the possible removal of DMUs the Andersen-Petersen measure can be regarded as deficient in its treatment of the nonzero slacks. It is also deficient because its treatment of the slacks does not yield a measure that is “units invariant.” We therefore turn to SBM (the slacks based measure) to eliminate these deficiencies. We will do this in a way that eliminates the non-solution possibilities that are present when the convexity condition $\sum_{j=1}^n \lambda_j = 1$ is adjoined to the models we will suggest.

10.3 NON-RADIAL SUPER-EFFICIENCY MODELS

Here, we discuss the super-efficiency issues under the assumption that the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is SBM-efficient, i.e., it is strongly efficient. (See Definition 4.6 in Chapter 4.)

Let us define a production possibility set $P \setminus (\mathbf{x}_o, \mathbf{y}_o)$ spanned by (X, Y) excluding $(\mathbf{x}_o, \mathbf{y}_o)$, i.e.

$$P \setminus (\mathbf{x}_o, \mathbf{y}_o) = \left\{ (\bar{\mathbf{x}}, \bar{\mathbf{y}}) \mid \bar{\mathbf{x}} \geq \sum_{j=1, \neq o}^n \lambda_j \mathbf{x}_j, \quad \bar{\mathbf{y}} \leq \sum_{j=1, \neq o}^n \lambda_j \mathbf{y}_j, \quad \bar{\mathbf{y}} \geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0} \right\}. \quad (10.2)$$

Further, we define a subset $\bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$ of $P \setminus (\mathbf{x}_o, \mathbf{y}_o)$ as

$$\bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o) = P \setminus (\mathbf{x}_o, \mathbf{y}_o) \cap \{\bar{\mathbf{x}} \geq \mathbf{x}_o \text{ and } \bar{\mathbf{y}} \leq \mathbf{y}_o\}, \quad (10.3)$$

where $P \setminus (\mathbf{x}_o, \mathbf{y}_o)$ means that point $(\mathbf{x}_o, \mathbf{y}_o)$ is excluded. By assumption $X > 0$ and $Y > 0$, $\bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$ is not empty.

As a weighted l_1 distance from $(\mathbf{x}_o, \mathbf{y}_o)$ and $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$, we employ the index δ as defined by

$$\delta = \frac{\frac{1}{m} \sum_{i=1}^m \bar{x}_i / x_{io}}{\frac{1}{s} \sum_{r=1}^s \bar{y}_r / y_{ro}}. \quad (10.4)$$

From (10.3), this distance is not less than 1 and attains 1 if and only if $(\mathbf{x}_o, \mathbf{y}_o) \in \bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$, i.e. exclusion of the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ has no effect on the original production possibility set P .

We can interpret this index as follows. The numerator is a weighted l_1 distance from \mathbf{x}_o to $\bar{\mathbf{x}} (\geq \mathbf{x}_o)$, and hence it expresses an average expansion rate of \mathbf{x}_o to $\bar{\mathbf{x}}$ of the point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$. The denominator is a weighted l_1 distance from \mathbf{y}_o to $\bar{\mathbf{y}} (\leq \mathbf{y}_o)$, and hence it is an average reduction rate of \mathbf{y}_o to $\bar{\mathbf{y}}$ of $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \bar{P} \setminus (\mathbf{x}_o, \mathbf{y}_o)$. The smaller the denominator is, the farther \mathbf{y}_o is positioned relative to $\bar{\mathbf{y}}$. Its inverse can be interpreted as an index of the distance from \mathbf{y}_o to $\bar{\mathbf{y}}$. Therefore, δ is a product of two indices: one, the distance in the input space, and the other in the output space. Both indices are dimensionless.

10.3.1 Definition of Non-radial Super-efficiency Measure

Based on the above observations, we define the super-efficiency of $(\mathbf{x}_o, \mathbf{y}_o)$ as the optimal objective function value δ^* from the following program:

$$\begin{aligned} [\text{SuperSBM-C}] \quad \delta^* &= \min_{\bar{\mathbf{x}}, \bar{\mathbf{y}}, \boldsymbol{\lambda}} \frac{\frac{1}{m} \sum_{i=1}^m \bar{x}_i / x_{io}}{\frac{1}{s} \sum_{r=1}^s \bar{y}_r / y_{ro}} \\ \text{subject to} \quad \bar{\mathbf{x}} &\geq \sum_{j=1, j \neq o}^n \lambda_j \mathbf{x}_j \\ \bar{\mathbf{y}} &\leq \sum_{j=1, j \neq o}^n \lambda_j \mathbf{y}_j \\ \bar{\mathbf{x}} &\geq \mathbf{x}_o \text{ and } \bar{\mathbf{y}} \leq \mathbf{y}_o \\ \bar{\mathbf{y}} &\geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \quad (10.5)$$

Let us introduce $\phi \in R^m$ and $\psi \in R^s$ such that

$$\bar{x}_i = x_{io}(1 + \phi_i) \quad (i = 1, \dots, m) \text{ and } \bar{y}_r = y_{ro}(1 + \psi_r) \quad (r = 1, \dots, s). \quad (10.6)$$

Then, this program can be equivalently stated in terms of ϕ , ψ and $\boldsymbol{\lambda}$ as follows:

$$\begin{aligned} [\text{SuperSBM-C}'] \quad \delta^* &= \min_{\phi, \psi, \boldsymbol{\lambda}} \frac{1 + \frac{1}{m} \sum_{i=1}^m \phi_i}{1 - \frac{1}{s} \sum_{r=1}^s \psi_r} \\ \text{subject to} \quad \sum_{j=1, j \neq o}^n x_{ij} \lambda_j - x_{io} \phi_i &\leq x_{io} \quad (i = 1, \dots, m) \\ \sum_{j=1, j \neq o}^n y_{rj} \lambda_j + y_{ro} \psi_r &\geq y_{ro} \quad (r = 1, \dots, s) \\ \phi_i &\geq 0 \ (\forall i), \quad \psi_r \geq 0 \ (\forall r), \quad \lambda_j \geq 0 \ (\forall j) \end{aligned} \quad (10.7)$$

We have the following two propositions (Tone (2002)⁶).

Proposition 10.1 *The super-efficiency score δ^* is units invariant, i.e. it is independent of the units in which the inputs and outputs are measured provided these units are the same for every DMU.*

Proof : This proposition holds, since both the objective function and constraints are units invariant. \square

Proposition 10.2 *Let $(\alpha \mathbf{x}_o, \beta \mathbf{y}_o)$ with $\alpha \leq 1$ and $\beta \geq 1$ be a DMU with reduced inputs and enlarged outputs than $(\mathbf{x}_o, \mathbf{y}_o)$. Then, the super-efficiency score of $(\alpha \mathbf{x}_o, \beta \mathbf{y}_o)$ is not less than that of $(\mathbf{x}_o, \mathbf{y}_o)$.*

Proof : The super-efficiency score ($\hat{\delta}^*$) of $(\alpha \mathbf{x}_o, \beta \mathbf{y}_o)$ is evaluated by solving the following program:

$$[\text{SuperSBM-C-2}] \quad \hat{\delta}^* = \min \frac{\frac{1}{m} \sum_{i=1}^m \hat{x}_i / (\alpha x_{io})}{\frac{1}{s} \sum_{r=1}^s \hat{y}_r / (\beta y_{ro})}$$

$$\begin{aligned}
&= \min \frac{\beta \frac{1}{m} \sum_{i=1}^m \hat{x}_i / x_{io}}{\alpha \frac{1}{s} \sum_{r=1}^s \hat{y}_r / y_{ro}} \\
\text{subject to} \quad &\hat{x} \geq \sum_{j=1, \neq o}^n \lambda_j x_j \\
&\hat{y} \leq \sum_{j=1, \neq o}^n \lambda_j y_j \\
&\hat{x} \geq \alpha x_o \text{ and } 0 \leq \hat{y} \leq \beta y_o \\
&\lambda \geq 0.
\end{aligned} \tag{10.8}$$

It can be observed that, for any feasible solution (\hat{x}, \hat{y}) for [SuperSBM-C-2], $(\hat{x}/\alpha, \hat{y}/\beta)$ is feasible for [SuperSBM-C]. Hence it holds

$$\delta^* \leq \frac{\frac{1}{m} \sum_{i=1}^m (\hat{x}_i/\alpha) / x_{io}}{\frac{1}{s} \sum_{r=1}^s (\hat{y}_r/\beta) / y_{ro}} = \frac{\beta \frac{1}{m} \sum_{i=1}^m \hat{x}_i / x_{io}}{\alpha \frac{1}{s} \sum_{r=1}^s \hat{y}_r / y_{ro}}. \tag{10.9}$$

Comparing (10.8) with (10.9) we see that:

$$\delta^* \leq \hat{\delta}^*.$$

Thus, the super-efficiency score of $(\alpha x_o, \beta y_o)$ ($\alpha \leq 1$ and $\beta \geq 1$) is not less than that of (x_o, y_o) . \square

10.3.2 Solving Super-efficiency

The fractional program [SuperSBM-C'] can be transformed into a linear programming problem using the Charnes-Cooper transformation as:

$$\begin{aligned}
[\text{LP}] \quad &\tau^* = \min t + \frac{1}{m} \sum_{i=1}^m \Phi_i \\
\text{subject to} \quad &t - \frac{1}{s} \sum_{r=1}^s \Psi_r = 1 \\
&\sum_{j=1, \neq o}^n x_{ij} \Lambda_j - x_{io} \Phi_i - x_{io} t \leq 0 \ (i = 1, \dots, m) \\
&\sum_{j=1, \neq o}^n y_{rj} \Lambda_j + y_{ro} \Psi_r - y_{ro} t \geq 0 \ (r = 1, \dots, s) \\
&\Phi_i \geq 0 \ (\forall i), \ \Psi_r \geq 0 \ (\forall r), \ \Lambda_j \geq 0 \ (\forall j)
\end{aligned} \tag{10.10}$$

Let an optimal solution of [LP] be $(\tau^*, \Phi^*, \Psi^*, \Lambda^*, t^*)$. Then we have an optimal solution of [SuperSBM-C'] expressed by

$$\delta^* = \tau^*, \ \lambda^* = \Lambda^*/t^*, \ \phi^* = \Phi^*/t^*, \ \psi^* = \Psi^*/t^*. \tag{10.11}$$

Furthermore, the optimal solution of [Super-SBM-C] is given by:

$$\bar{x}_{io}^* = x_{io}(1 + \phi_i^*) \ (i = 1, \dots, m) \text{ and } \bar{y}_{ro}^* = y_{ro}(1 - \psi_r^*) \ (r = 1, \dots, s). \tag{10.12}$$

10.3.3 Input/Output-Oriented Super-efficiency

In order to adapt the non-radial super-efficiency model to input (output) orientation, we can modify the preceding program as follows. For input orientation, we deal with the weighted l_1 -distance only in the input space. Thus, the program turns out to be:

$$\begin{aligned} [\text{SuperSBM-I-C}] \quad \delta_I^* &= \min_{\phi, \lambda} 1 + \frac{1}{m} \sum_{i=1}^m \phi_i && (10.13) \\ \text{subject to} \quad & \sum_{j=1, \neq o}^n x_{ij} \lambda_j - x_{io} \phi_i \leq x_{io} \quad (i = 1, \dots, m) \\ & \sum_{j=1, \neq o}^n y_{rj} \lambda_j \geq y_{ro} \quad (r = 1, \dots, s) \\ & \phi_i \geq 0 \ (\forall i), \ \lambda_j \geq 0 \ (\forall j) \end{aligned}$$

In a similar way we can develop the output-oriented super-efficiency model as follows:

$$\begin{aligned} [\text{SuperSBM-O-C}] \quad \delta_O^* &= \min_{\psi, \lambda} \frac{1}{1 - \frac{1}{s} \sum_{r=1}^s \psi_r} && (10.14) \\ \text{subject to} \quad & \sum_{j=1, \neq o}^n x_{ij} \lambda_j \leq x_{io} \quad (i = 1, \dots, m) \\ & \sum_{j=1, \neq o}^n y_{rj} \lambda_j + y_{ro} \psi_r \geq y_{ro} \quad (r = 1, \dots, s) \\ & \psi_r \geq 0 \ (\forall r), \ \lambda_j \geq 0 \ (\forall j) \end{aligned}$$

Since the above two models have the same restricted feasible region as [SuperSBM-C], we have:

Proposition 10.3 $\delta_I^* \geq \delta^*$ and $\delta_O^* \geq \delta^*$, where δ^* is defined in (10.4).

10.3.4 An Example of Non-radial Super-efficiency

Using the [SuperSBM-I-C] model in (10.13), we solved the same data set described in Table 10.1 and obtained the results exhibited in Table 10.3. Since DMUs A and E are enveloped respectively by B and C, their SBM scores are less than unity and SBM-inefficient. Compared with the radial model, DMU A dropped its efficiency from $1 - 4\epsilon$ to 0.833. This is caused by the slack $s_2^- = 4$. We need to solve [SuperSBM-I-C] for DMUs B, C, D and F which have SBM score unity. DMU B has the optimal solution $\lambda_A^* = 1$, $s_2^* = x_{2A} \phi_2^* = 4$, ($\phi_2^* = 4/8 = 0.5$) with all other variables zero, and hence its super-efficiency score is $1 + 0.5/2 = 1.25$. Similarly, DMUs C, D and F have super-efficiency scores 1.092, 1.125 and 1, respectively. In this example, the

ranking is the same as the Andersen-Petersen model, although they are not always same (see Tone (2002)).

Table 10.3. Non-radial Super-efficiency

DMU	A	B	C	D	E	F
Super-eff.	0.833	1.25	1.092	1.125	0.667	1
Rank	5	1	3	2	6	4
Reference set	$\lambda_B = 1$	$\lambda_A = 1$	$\lambda_D = 0.23$ $\lambda_F = 0.77$	$\lambda_C = 1$	$\lambda_C = 1$	$\lambda_B = 0.5$ $\lambda_C = 0.5$
Slacks						
$s_1^- = x_1\phi_1$	0	0	0	0	5	0
$s_2^- = x_2\phi_2$	4	4	0.923	1	1	0

10.4 EXTENSIONS TO VARIABLE RETURNS-TO-SCALE

In this section, we extend our super-efficiency models to the variable returns-to-scale models and discuss the infeasible LP issues that can then arise.

We extend our analysis to the variable returns-to-scale case by adjoining the following convexity constraints to the models:

$$\sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0, \quad \forall j. \quad (10.15)$$

We provide two approaches as follows:

10.4.1 Radial Super-efficiency Case

We have two models [SuperRadial-I-V] (Input-oriented Variable RTS) and [SuperRadial-O-V] (Output-oriented Variable RTS) with models represented respectively as follows.

$$[\text{SuperRadial-I-V}] \quad \theta^* = \min \theta \quad (10.16)$$

$$\text{subject to} \quad \theta x_{io} \geq \sum_{j=1, \neq o}^n \lambda_j x_{ij} \quad (i = 1, \dots, m) \quad (10.17)$$

$$y_{ro} \leq \sum_{j=1, \neq o}^n \lambda_j y_{rj} \quad (r = 1, \dots, s) \quad (10.18)$$

$$\begin{aligned} & \sum_{j=1, \neq o}^n \lambda_j = 1 \\ & \lambda_j \geq 0 \quad (\forall j). \end{aligned} \quad (10.19)$$

$$[\text{SuperRadial-O-V}] \quad 1/\eta^* = \min 1/\eta \quad (10.20)$$

$$\text{subject to} \quad x_{io} \geq \sum_{j=1, \neq o}^n \lambda_j x_{ij} \quad (i = 1, \dots, m) \quad (10.21)$$

$$\eta y_{ro} \leq \sum_{j=1, \neq o}^n \lambda_j y_{rj} \quad (r = 1, \dots, s) \quad (10.22)$$

$$\begin{aligned} & \sum_{j=1, \neq o}^n \lambda_j = 1 \\ & \lambda_j \geq 0 \quad (\forall j). \end{aligned} \quad (10.23)$$

The above two programs may suffer from infeasibility under the following conditions. Suppose, for example, y_{1o} is larger than the other y_{1j} ($j \neq o$), i.e.,

$$y_{1o} > \max_{j=1, \neq o} \{y_{1j}\}.$$

Then, the constraint (10.18) in [SuperRadial-I-V] is infeasible for $r = 1$ by dint of the constraint (10.19).

Likewise, suppose, for example, x_{1o} is smaller than the other x_{1j} ($j \neq o$), i.e.,

$$x_{1o} < \min_{j=1, \neq o} \{x_{1j}\}.$$

Then, the constraint (10.21) in [SuperRadial-O-V] is infeasible for $i = 1$ by dint of the constraint (10.23).

Thus, we have:

Proposition 10.4 [SuperRadial-I-V] *has no feasible solution if there exists r such that $y_{ro} > \max_{j \neq o} \{y_{rj}\}$, and [SuperRadial-O-V] has no feasible solution if there exists i such that $x_{io} < \min_{j \neq o} \{x_{ij}\}$.*

We notice that Proposition 10.4 above is not a necessary condition for infeasibility, i.e., infeasibility may occur in other cases.

10.4.2 Non-radial Super-efficiency Case

The non-radial super-efficiency under variable returns-to-scale is evaluated by solving the following program:

$$[\text{SuperSBM-V}] \quad \delta^* = \min \delta = \frac{\frac{1}{m} \sum_{i=1}^m \bar{x}_i / x_{io}}{\frac{1}{s} \sum_{r=1}^s \bar{y}_r / y_{ro}} \quad (10.24)$$

$$\text{subject to} \quad \bar{x} \geq \sum_{j=1, \neq o}^n \lambda_j x_j$$

$$\bar{y} \leq \sum_{j=1, \neq o}^n \lambda_j y_j$$

$$\begin{aligned}\bar{x} &\geq x_o \text{ and } \bar{y} \leq y_o \\ \sum_{j=1, \neq o}^n \lambda_j &= 1 \\ \bar{y} &\geq 0, \quad \lambda \geq 0.\end{aligned}$$

Under the assumptions $X > O$ and $Y > O$, we demonstrate that the [SuperSBM-V] is always feasible and has a finite optimum in contrast to the radial super-efficiency. This can be shown as follows.

We choose a DMU $j(\neq o)$ with (x_j, y_j) and set $\bar{\lambda}_j = 1$ and $\bar{\lambda}_k = 0$ ($k \neq j$). Using this DMU (x_j, y_j) we define:

$$\tilde{x}_i = \max\{x_{io}, x_{ij}\} \quad (i = 1, \dots, m) \quad (10.25)$$

$$\tilde{y}_r = \min\{y_{ro}, y_{rj}\} \quad (r = 1, \dots, s). \quad (10.26)$$

Thus, the set $(\bar{x} = \tilde{x}, \bar{y} = \tilde{y}, \lambda = \bar{\lambda})$ is feasible for the [SuperSBM-V]. Hence, [SuperSBM-V] is always feasible with a finite optimum. Thus, we have:

Theorem 10.1 (Tone (2002)) *The non-radial super-efficiency model under the variable returns-to-scale environment,[SuperSBM-V], is always feasible and has a finite optimum.*

We can define [SuperSBM-I-V] (Input-oriented Variable RTS) and [SuperSBM-O-V] (Output-oriented Variable RTS) models similar to [SuperSBM-I-C] and [SuperSBM-O-C]. We notice that [SuperSBM-I-V] and [SuperSBM-O-V] models confront the same infeasible LP issues as the [SuperRadial-I-V] and [Super Radial-O-V]. See Problem 10.1 for comparisons of super-efficiency models.

10.5 SUMMARY OF CHAPTER 10

This chapter introduced the concept of super-efficiency and presented two types of approach for measuring super-efficiency: radial and non-radial. Super-efficiency measures are widely utilized in DEA applications for many purposes, e.g., ranking efficient DMUs, evaluating the Malmquist productivity index and comparing performances of two groups (the bilateral comparisons model in Chapter 7).

10.6 NOTES AND SELECTED BIBLIOGRAPHY

Andersen and Petersen (1993) introduced the first super-efficiency model for radial models. Tone (2002) introduced non-radial super-efficiency models using the SBM.

10.7 RELATED DEA-SOLVER MODELS FOR CHAPTER 10

Super-efficiency codes have the same data format with the CCR model.

Super-Radial-I(O)-C(V) (Input(Output)-oriented Radial Super-efficiency

model).

This code solves the oriented radial super-efficiency model under the constant (variable) returns-to-scale assumption. If the corresponding LP for a DMU is infeasible, we return a score 1 to the DMU.

Super-SBM-I(O)-C(V or GRS) (Input(Output)-oriented Non-radial Super-efficiency model)

This code solves the oriented non-radial super-efficiency model under the constant (variable or general) returns-to-scale assumption. If the corresponding LP for a DMU is infeasible, we return a score 1 to the DMU.

Super-SBM-C(V or GRS) (Non-radial and Non-oriented Super-efficiency model)

This code solves the non-oriented non-radial super-efficiency model under the constant (variable or general) returns-to-scale assumption.

10.8 PROBLEM SUPPLEMENT FOR CHAPTER 10

Problem 10.1

Table 10.4 displays data for 6 DMUs (A, B, C, D, E, F) with two inputs (x_1, x_2) and two outputs (y_1, y_2). Obtain and compare the super-efficiency scores using the attached DEA-Solver.

Table 10.4. Data for Super-efficiency

DMU	A	B	C	D	E	F
Input 1 (x_1)	2	2	5	10	10	3.5
Input 2 (x_2)	12	8	5	4	6	6.5
Output 1 (y_1)	4	3	2	2	1	1
Output 2 (y_2)	1	1	1	1	1	1

Suggested Answer : Tables 10.5 and 10.6 exhibit super-efficiency scores respectively under variable and constant returns-to-scale conditions. In the tables I (O) indicates the Input (Output) orientation and V (C) denotes the Variable (Constant) RTS. “NA” (not available) represents an occurrence of infeasible LP solution.

It is observed that, by dint of Proposition 10.4, Super-BCC-I and Super-SBM-I-V scores of DMU A are not available since it has $y_{1A} = 4$ which is strictly larger than other y_{1j} ($j \neq A$), and Super-BCC-O and Super-SBM-O-V scores of DMU D are not available since it has $x_{2D} = 4$ which is strictly smaller than other x_{2j} ($j \neq D$). The infeasibility of DMUs B and C for Super-BCC-O and Super-SBM-O-V is caused simply by non-existence of feasible solution for

the corresponding linear programs. However, Super-SBM-V always has finite solutions as claimed by Theorem 10.1.

Under the constant RTS assumption, all DMUs have a finite score which results from exclusion of the convexity constraint on λ_j . The differences in scores between radial and non-radial models are due to non-zero slacks.

Table 10.5. Super-efficiency Scores under Variable RTS

DMU	A	B	C	D	E	F
Super-BCC-I	NA	1.26	1.17	1.25	0.75	1
Super-BCC-O	1.33	NA	NA	NA	1	1
Super-SBM-I-V	NA	1.25	1.12	1.13	0.66	1
Super-SBM-O-V	1.14	NA	NA	NA	0.57	0.57
Super-SBM-V	1.14	1.25	1.12	1.13	0.42	0.57

Table 10.6. Super-efficiency Scores under Constant RTS

DMU	A	B	C	D	E	F
Super-CCR-I	1.33	1.26	1.17	1.25	0.75	1
Super-CCC-O	1.33	1.26	1.17	1.25	0.75	1
Super-SBM-I-C	1.17	1.25	1.12	1.13	0.67	1
Super-SBM-O-C	1.14	1.2	1.09	1.25	0.5	0.57
Super-SBM-C	1.14	1.16	1.09	1.13	0.42	0.57

Notes

1. P. Andersen and N.C. Petersen (1993), "A Procedure for Ranking Efficient Units in Data Envelopment Analysis," *Management Science* 39, pp.1261-1264.
2. P. Wilson (1993), "Detecting Influential Observations in Data Envelopment Analysis," *Journal of Productivity Analysis* 6, pp. 27-46.
3. S. Ray (2004), *Data Envelopment Analysis* (Cambridge: Cambridge University Press).
4. A. Charnes, W.W. Cooper and R.M. Thrall (1991), "A Structure for Classifying and Characterizing Inefficiency in DEA," *Journal of Productivity Analysis* 2, pp. 197-237.
5. Andersen and Petersen report this as $\theta^* = 0.8333$, but their reported value is seen to be incorrect since $0.8333 > 0.75$.
6. K. Tone (2002), "A Slacks-based Measure of Super-efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 143, pp.32-41

11 EFFICIENCY CHANGE OVER TIME

11.1 INTRODUCTION

As just indicated, most of this text has dealt with uses of DEA under static conditions. This can be misleading since dynamic settings may give rise to seemingly excessive use of resources which are intended to produce beneficial results in future periods. This chapter presents non-parametric DEA models under dynamic (time dependent) situations.

We first introduce a time-dependent use of DEA that is known as “window analysis.” This name and the basic concept are due to G. Klopp (1985)¹ who developed these techniques in his capacity as chief statistician for the U.S. Army Recruiting Command.

Then, we present the “Malmquist productivity index.” The concept of this index was first introduced by S. Malmquist (1953),² and has further been studied and developed in the non-parametric framework by several authors. See for example, among others, Caves *et al.* (1982),³ Färe *et al.* (1992),⁴ Färe *et al.* (1989, 1994),⁵ Färe *et al.* (1998b),⁶ and Thrall (2000).⁷ It is an index representing Total Factor Productivity (TFP) growth of a Decision Making Unit (DMU), in that it reflects (1) progress or regress in efficiency along with (2) progress or regress of the frontier technology between two periods of time under the multiple inputs and multiple outputs framework.

11.2 WINDOW ANALYSIS

11.2.1 An Example

Although now used in many other contexts we here revert to the applications in army recruiting efforts that gave rise to “window analysis.” This example is built around 3 outputs consisting of recruits such as male and female high school graduates and degree candidates (such as high school seniors not yet graduated). The 10 inputs consist of number of recruiters (such as recruiting sergeants) and their aids, amount spent on local advertising, as well as qualities such as recruiter experience which are all discretionary variables as well as non-discretionary variables such as local unemployment, number of high school seniors in the local market and their propensity to enlist.⁸

These variables are used (among others) by the U.S. Army Recruiting Command (USAREC) to evaluate performance in its various organization units which Klopp (1985, p.115) describes as follows. USAREC recruits for the entire United States. To facilitate control of the recruiting process USAREC divides the U.S. into 5 regions managed by entities referred to as “Recruiting Brigades.” Each Recruiting Brigade is responsible for some of the 56 “Recruiting Battalions” that operate in the U.S. and the latter are responsible, in turn, for the “Recruiting Stations” where recruiters are assigned specific missions. For the example we employ, it was decided to use the Recruiting Battalions each of which was designated as a DMU.

The DEA analyses took a variety of forms. One consisted of standard (static) DEA reports in the manner of Table 3.7 expanded to include detailed information on the inefficiencies present in each DMU. Both discretionary and non-discretionary variables were included along with the members of the peer group used to effect the evaluation.

Something more was wanted in the form of trend analyses of the quarterly reports that USAREC received. A use of statistical regressions and time series analyses of the efficiency scores proved unsatisfactory. Experimentation was therefore undertaken, which led to the window analysis we now describe.

11.2.2 Application

Table 11.1, as adapted from Klopp (1985) will be used for guidance in this discussion. The basic idea is to regard each DMU as if it were a different DMU in each of the reporting dates represented by Q1, Q2, etc., at the top of the table. For instance, the results in row 1 for Battalion 1A represent four values obtained by using its results in Q1, Q2, Q3 and Q4 by bringing this DMU into the objective for each of these quarters. Because these 56 DMUs are regarded as different DMUs in each quarter, these evaluations are conducted by reference to the entire set of $4 \times 56 = 224$ DMUs that are used to form the data matrix. Thus the values of $\theta^* = 0.83, 1.00, 0.95$ and 1.00 in row 1 for Battalion 1A represent its quarterly performance ratings as obtained from this matrix with $13 \times 224 = 2,912$ entries.

Table 11.1. Window Analysis: 56 DMUs in U.S. Army Recruitment Battalions 3 Outputs - 10 Inputs

Battalion	Efficiency Scores								Summary		Measures	
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Mean	Var	Column Range	Total Range
1A	0.83	1.00	0.95	1.00								
		1.00	0.95	1.00	1.00							
			1.00	1.00	1.00	1.00						
				1.00	1.00	1.00	1.00					
					1.00	1.00	1.00	1.00				
						1.00	1.00	1.00				
							1.00	1.00				
								1.00				
									0.99	0.03	.05	.17
1B	0.71	1.00	0.91	0.89								
		1.00	0.91	0.89	0.77							
			1.00	1.00	0.76	1.00						
				1.00	0.75	1.00	1.00					
					0.80	1.00	1.00	1.00				
						1.00	1.00	1.00				
1C	0.72	1.00	0.86	1.00								
		1.00	0.87	1.00	0.74							
			0.87	1.00	0.74	1.00						
				1.00	1.00	1.00	1.00					
					0.80	1.00	1.00	1.00				
						1.00	1.00	1.00				
1D	0.73	1.00	1.00	1.00								
		1.00	1.00	1.00	1.00							
			1.00	1.00	1.00	1.00						
				1.00	1.00	1.00	1.00					
					0.80	1.00	1.00	1.00				
						1.00	1.00	1.00				
1E	0.69	1.00	0.83	1.00								
		0.83	0.83	1.00	0.77							
			0.83	1.00	0.79	0.93						
				1.00	0.77	0.95	0.83					
					0.82	0.88	0.88	1.00				
1F	1.00	1.00	1.00	1.00								
		1.00	1.00	1.00	1.00							
			1.00	1.00	1.00	1.00						
				1.00	1.00	1.00	1.00					
					0.80	1.00	1.00	1.00				
1G	0.65	0.73	0.68	0.77								
		0.73	0.68	0.77	0.68							
			0.68	0.77	0.68	0.72						
				0.76	0.74	0.75	1.00					
					1.00	1.00	1.00	1.00				
1H	0.82	1.00	1.00	1.00								
		1.00	1.00	1.00	0.71							
			1.00	1.00	0.71	0.70						
				1.00	1.00	0.72	0.72					
					0.79	0.81	1.00	1.00				
1I	1.00	1.00	1.00	1.00								
		1.00	1.00	1.00	1.00							
			0.86	1.00	1.00	1.00						
				1.00	1.00	1.00	1.00					
					1.00	1.00	1.00	1.00				
1K	0.81	1.00	1.00	1.00								
		1.00	1.00	1.00	0.73							
			1.00	1.00	0.73	0.67						
				1.00	0.85	0.67	0.67					
					0.67	0.67	0.67	0.67				

Similar first-row results for other DMUs are obtained for each of the 10 DMUs in Table 11.1, which we have extracted from a larger tabulation of 56 DMUs to obtain compactness. After first row values have been similarly obtained, a new 4-period window is obtained by dropping the data for Q1 and adding the data for Q5. Implementing the same procedure as before produces the efficiency ratings in row 2 for each Battalion. The process is then continued until no further quarters are added — as occurs (here) in “row” 5 with its 4 entries.

To help interpret the results we note that the “column views” enable us to examine the stability of results across the different data sets that occur with these removal and replacement procedures. “Row views” make it possible to determine trends and/or observed behavior with the same data set. Thus, the column view for Battalion 1A shows stability and the row view shows steady behavior after the improvement over Q1. At the bottom of Table 11.1, however, Battalion 1K exhibits deteriorating behavior and this same deterioration continues to be manifested with different data sets. Moreover, to reinforce this finding the values in each column for Battalion 1K do not change very much and this stability reinforces this finding.

Finally we augment this information by the statistical values noted on the right side of Table 11.1. Here the average is used as a representative measure obtained from the θ^* values for each DMU and matched against its variance. Medians might also be used and matched with ranges and so on. Other summary statistics and further decompositions may also be used, of course, but a good deal of information is supplied in any case.

Weaknesses are also apparent, of course, such as the absence of attention to nonzero slacks. However, the principle and formulas to be supplied in the next section of this chapter may be applied to slack portrayals, too, if desired and other similar measures may be used such as the SBM (Slacks Based Measure) given in Section 4.4 of Chapter 4.

Another deficiency is apparent in that the beginning and ending period DMUs are not tested as frequently as the others. Thus the DMUs in Q1 are examined in only one window and the same is true for Q8. In an effort to address this problem Sueyoshi (1992)⁹ introduced a “round robin” procedure which proceeds as follows: First each period is examined independently. This is followed by a 2-period analysis after which a three-period analysis is used. And so on. However, this analysis becomes unwieldy since the number of combinations grows to $2^p - 1$ so that some care is needed with this approach.

11.2.3 Analysis

The following formulas adapted from D.B. Sun (1988)¹⁰ can be used to study the properties of these window analyses. For this purpose we introduce the following symbols

$$\begin{aligned} n &= \text{number of DMUs} \\ k &= \text{number of periods} \end{aligned} \tag{11.1}$$

p = length of window ($p \leq k$)

w = number of windows.

We then reduce the number of DMUs to the 10 in Table 11.1 so we can use it for numerical illustrations.

	Formula	Application (to Table 11.1)
no. of windows:	$w = k - p + 1$	$8 - 4 + 1 = 5$
no. of DMUs in each window:	$np/2$	$10 \times 4/2 = 20$
no. of "different" DMUs:	npw	$10 \times 4 \times 5 = 200$
Δ no. of DMUs:	$n(p-1)(k-p)$	$10 \times 3 \times 4 = 120$

Here " Δ " represents an increase compared to the $8 \times 10 = 80$ DMUs that would have been available if the evaluation had been separately effected for each of the 10 DMUs in each of the 8 quarters.

An alternate formula for deriving the total number of DMUs is given in Charnes and Cooper (1991)¹¹ as follows.

$$\text{Total no. of "different" DMUs : } n(k-p+1)p = 10 \times (8-4+1) \times 4 = 200 \quad (11.2)$$

Differentiating this last function with respect to p and equating to zero gives

$$p = \frac{k+1}{2} \quad (11.3)$$

as the condition for a maximum number of DMUs. This result need not be an integer, however, so we utilize the symmetry of (11.2) and (11.3) and modify the latter to the following,

$$p = \begin{cases} \frac{k+1}{2} & \text{when } k \text{ is odd} \\ \frac{k+1}{2} \pm \frac{1}{2} & \text{when } k \text{ is even.} \end{cases} \quad (11.4)$$

To see how to apply this formula when k is even we first note that

$$n(k-p+1)p = n [(k+1)p - p^2].$$

Hence by direct substitution we obtain

$$\begin{aligned} & n \left[(k+1) \left(\frac{k+1}{2} - \frac{1}{2} \right) - \left(\frac{k+1}{2} - \frac{1}{2} \right)^2 \right] \\ &= n \left[(k+1) \left(\frac{k+1}{2} + \frac{1}{2} \right) - \left(\frac{k+1}{2} + \frac{1}{2} \right)^2 \right] = \frac{n}{4} [(k+1)^2 - 1] \end{aligned}$$

Then, for $k = 8$, as in our example, we find from (11.3) that

$$p = \frac{8+1}{2} = 4.5$$

which is not an integer. Hence using $[p]$ to mean “the integer closest to p we apply the bottom expression in (11.4) to obtain

$$[p] = \begin{cases} 4 = 4.5 - 0.5 \\ 5 = 4.5 + 0.5 \end{cases}$$

and note that substitution in (11.2) produces 200 “different” DMUs as the maximum number in either case.

11.3 MALMQUIST INDEX

In this section we first introduce a concept of Malmquist productivity index, and then present three different approaches for its measurement: radial, non-radial, and non-radial and non-oriented approaches. Scale efficiency change is discussed in Section 11.3.8. We demonstrate with numerical examples for comparing different approaches in Section 11.3.9. Concluding remarks follow in Section 11.3.10.

11.3.1 Dealing with Panel Data

The Malmquist index evaluates the productivity change of a DMU between two time periods and is an example in “comparative statics” analysis. It is defined as the product of *Catch-up* and *Frontier-shift* terms. The catch-up (or recovery) term relates to the degree to which a DMU improves or worsens its efficiency, while the frontier-shift (or innovation) term reflects the change in the efficient frontiers between the two time periods. We deal with a set of n DMUs $(\mathbf{x}_j, \mathbf{y}_j)$ ($j = 1, \dots, n$) each having m inputs denoted by a vector $\mathbf{x}_j \in R^m$ and q outputs denoted by a vector $\mathbf{y}_j \in R^q$ over the periods 1 and 2. We assume $\mathbf{x}_j > \mathbf{0}$ and $\mathbf{y}_j > \mathbf{0}$ ($\forall j$). The notations $(\mathbf{x}_o, \mathbf{y}_o)^1 = (\mathbf{x}_o^1, \mathbf{y}_o^1)$ and $(\mathbf{x}_o, \mathbf{y}_o)^2 = (\mathbf{x}_o^2, \mathbf{y}_o^2)$ are employed for designating DMU _{o} ($o = 1, \dots, n$) in periods 1 and 2 respectively. The production possibility set $(X, Y)^t$ ($t = 1$ and 2) spanned by $(\mathbf{x}_j, \mathbf{y}_j)^t$ ($j = 1, \dots, n$) is defined by:

$$(X, Y)^t = \left\{ (\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \sum_{j=1}^n \lambda_j \mathbf{x}_j^t, \mathbf{0} \leq \mathbf{y} \leq \sum_{j=1}^n \lambda_j \mathbf{y}_j^t, \mathbf{L} \leq \mathbf{e}\boldsymbol{\lambda} \leq \mathbf{U}, \boldsymbol{\lambda} \geq \mathbf{0} \right\}, \quad (11.5)$$

where \mathbf{e} is the row vector with all elements equal to one, $\boldsymbol{\lambda} \in R^n$ is the intensity vector, and L and U are the lower and upper bounds for the sum of the intensities. $(L, U) = \{(0, \infty), (1, 1), (1, \infty) \text{ and } (0, 1)\}$ correspond to the CCR (Constant Returns to Scale = CRS), BCC (Variable Returns to Scale = VRS), IRS (Increasing Returns to Scale) and DRS (Decreasing Returns to Scale) models. The production possibility set $(X, Y)^t$ is characterized by frontiers that are composed of $(\mathbf{x}, \mathbf{y}) \in (X, Y)^t$ such that it is not possible to improve any element of the input \mathbf{x} or any element of the output \mathbf{y} without worsening some other input or output. We call this frontier set the frontier technology at period t . In the Malmquist index analysis, the efficiencies of DMUs $(\mathbf{x}_o, \mathbf{y}_o)^1$ and $(\mathbf{x}_o, \mathbf{y}_o)^2$ are evaluated by the frontier technologies 1 and 2 in several ways.

11.3.2 Catch-up Effect

The catch-up effect from period 1 to 2 is measured by the following formula.

$$\text{Catch-up} = \frac{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^2 \text{ with respect to period 2 frontier}}{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^1 \text{ with respect to period 1 frontier}}. \quad (11.6)$$

We evaluate catch-up of each DMU defined in the above formula by appropriate DEA models. A simple example in the case of a single input and output technology is illustrated in Figure 11.1.

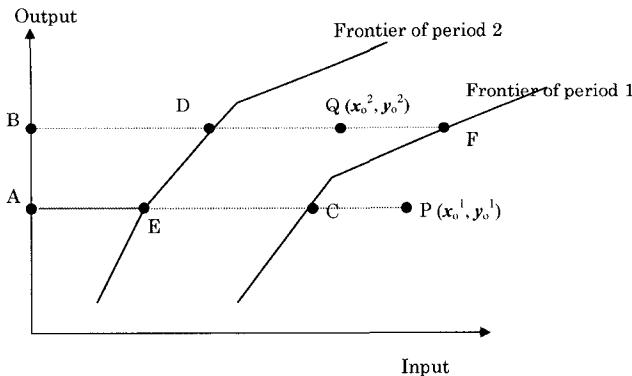


Figure 11.1. Catch-up

The catch-up effect (in an input-orientation) can be computed as:

$$\text{Catch-up} = \frac{BD}{BQ} / \frac{AC}{AP}. \quad (11.7)$$

$(\text{Catch-up}) > 1$ indicates progress in relative efficiency from period 1 to 2, while $(\text{Catch-up}) = 1$ and $(\text{Catch-up}) < 1$ respectively indicate no change and regress in efficiency.

11.3.3 Frontier-shift Effect

In addition to the catch-up term, we must take account of the frontier-shift (innovation) effect in order to fully evaluate the productivity change since the catch-up effect is determined by the efficiencies being measured by the distances from the respective frontiers. In Figure 11.1 case, this can be explained as follows. The reference point C of $(\mathbf{x}_o, \mathbf{y}_o)^1$ moved to E on the frontier of period 2. Thus, the frontier-shift effect at $(\mathbf{x}_o, \mathbf{y}_o)^1$ is evaluated by

$$\phi_1 = \frac{AC}{AE}. \quad (11.8)$$

This is equivalent to

$$\phi_1 = \frac{\frac{AC}{AP}}{\frac{AE}{AP}} = \frac{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^1 \text{ wrt period 1 frontier}}{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^1 \text{ wrt period 2 frontier}}. \quad (11.9)$$

The numerator of the right hand side of (11.9) is already available in (11.6). The denominator in (11.9) is measured as the efficiency score of $(\mathbf{x}_o, \mathbf{y}_o)^1$ relative to the period 2 frontier. Similarly, the frontier-shift effect at $(\mathbf{x}_o, \mathbf{y}_o)^2$ is expressed by

$$\phi_2 = \frac{\frac{BF}{BQ}}{\frac{BD}{BQ}} = \frac{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^2 \text{ wrt period 1 frontier}}{\text{Efficiency of } (\mathbf{x}_o, \mathbf{y}_o)^2 \text{ wrt period 2 frontier}}. \quad (11.10)$$

Using ϕ_1 and ϕ_2 , we define *Frontier-shift* effect by their geometric mean, i.e.,

$$\text{Frontier-shift} = \phi = \sqrt{\phi_1 \phi_2}, \quad (11.11)$$

where $\phi_1 \phi_2 = \frac{AC}{AE} \frac{BF}{BD}$.

$(\text{Frontier-shift}) > 1$ indicates progress in the frontier technology around DMU_o from period 1 to 2, while $(\text{Frontier-shift}) = 1$ and $(\text{Frontier-shift}) < 1$ respectively indicate the status quo and regress in the frontier technology.

11.3.4 Malmquist Index

The Malmquist index (MI) is computed as the product of (Catch-up) and (Frontier-shift), i.e.,

$$MI = (\text{Catch-up}) \times (\text{Frontier-shift}). \quad (11.12)$$

Thus combining (11.9) and (11.10) with (11.7) we get

$$MI = \frac{AP}{BQ} \sqrt{\frac{BF}{AC} \frac{BD}{AE}},$$

where the first term represents the relative change in performance and the second represents the relative change in the frontier used to evaluate these performances.

We now develop the numerical measures for which we employ the following notation for the efficiency score of DMU $(\mathbf{x}_o, \mathbf{y}_o)^{t_1}$ measured by the frontier technology t_2 .

$$\delta^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1}) \quad (t_1 = 1, 2 \text{ and } t_2 = 1, 2). \quad (11.13)$$

Using this notation, the catch-up effect (C) in (11.6) can be expressed as

$$C = \frac{\delta^2((\mathbf{x}_o, \mathbf{y}_o)^2)}{\delta^1((\mathbf{x}_o, \mathbf{y}_o)^1)}. \quad (11.14)$$

The frontier-shift effect, F, in (11.11) is represented by

$$F = \left[\frac{\delta^1((x_o, y_o)^1)}{\delta^2((x_o, y_o)^1)} \times \frac{\delta^1((x_o, y_o)^2)}{\delta^2((x_o, y_o)^2)} \right]^{1/2}. \quad (11.15)$$

As the product of C and F, we obtain the following formula for the computation of MI, the Malmquist Index,

$$MI = \left[\frac{\delta^1((x_o, y_o)^2)}{\delta^1((x_o, y_o)^1)} \times \frac{\delta^2((x_o, y_o)^2)}{\delta^2((x_o, y_o)^1)} \right]^{1/2}. \quad (11.16)$$

This last expression gives another interpretation of MI as the geometric mean of the two efficiency ratios: the one being the efficiency change measured by the period 1 technology and the other the efficiency change measured by the period 2 technology. As can be seen from these formulas, MI consists of four terms: $\delta^1((x_o, y_o)^1)$, $\delta^2((x_o, y_o)^2)$, $\delta^1((x_o, y_o)^2)$ and $\delta^2((x_o, y_o)^1)$. The first two are related to the measurements within the same time period with $t = 1$ or $t = 2$, while the last two are for intertemporal comparison. $MI > 1$ indicates progress in the total factor productivity of the DMU_o from period 1 to 2, while $MI = 1$ and $MI < 1$ respectively indicate the status quo and deterioration in the total factor productivity.

In a non-parametric framework the Malmquist index (MI) is constructed by means of DEA technologies. There are a number of ways to compute MI. First, Färe, Grosskopf, Lindgren and Roos (1989, 1994) utilized the input and output oriented radial DEA model to compute MI. However, the radial model suffers from one shortcoming, i.e., neglect of slacks. To overcome this shortcoming, MI can be computed using the *slack-based* non-radial and oriented DEA model. Alternatively MI can be computed from the non-radial and non-oriented DEA model.

11.3.5 The Radial MI

We start with the input and output oriented radial models. The input-oriented radial MI measures the within and intertemporal scores by the 4 linear programs given below.

$$\begin{aligned} & [\text{Within score in input-orientation}] \\ & \delta^s((x_o, y_o)^s) = \min_{\theta, \lambda} \theta \\ & \text{subject to} \\ & \theta x_o^s \geq X^s \lambda \\ & y_o^s \leq Y^s \lambda \\ & L \leq e \lambda \leq U \\ & \lambda \geq 0, \end{aligned} \quad (11.17)$$

where $X^s = (\mathbf{x}_1^s, \dots, \mathbf{x}_n^s)$ and $Y^s = (\mathbf{y}_1^s, \dots, \mathbf{y}_n^s)$ are respectively input and output matrices of observed data for period s . We solve this program for $s = 1$ and 2. It holds that $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^s) \leq 1$, and $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^s) = 1$ indicates $(\mathbf{x}_o, \mathbf{y}_o)^s$ being on the technically efficient frontiers of $(X, Y)^s$.

[Intertemporal score in input-orientation]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\theta, \boldsymbol{\lambda}} \theta \quad (11.18)$$

subject to

$$\theta \mathbf{x}_o^t \geq X^s \boldsymbol{\lambda}$$

$$\mathbf{y}_o^t \leq Y^s \boldsymbol{\lambda}$$

$$L \leq e\boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}.$$

We solve this program for the pairs $(s, t) = (1, 2)$ and $(2, 1)$. If $(\mathbf{x}_o, \mathbf{y}_o)^t$ is not enveloped by the technology of period s , the score $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t)$, if it exists, results in a value greater than 1. This corresponds to the concept of super-efficiency proposed by Andersen and Petersen (1993).¹² Refer to Chapter 10 for details. The above schemes are input-oriented. We can develop the output-oriented MI as well by means of the output-oriented radial DEA models. This is explained below.

[Within score in output-orientation]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^s) = \min_{\theta, \boldsymbol{\lambda}} \theta \quad (11.19)$$

subject to

$$\mathbf{x}_o^s \geq X^s \boldsymbol{\lambda}$$

$$(1/\theta)\mathbf{y}_o^s \leq Y^s \boldsymbol{\lambda}$$

$$L \leq e\boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}.$$

[Intertemporal score in output-orientation]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\theta, \boldsymbol{\lambda}} \theta \quad (11.20)$$

subject to

$$\mathbf{x}_o^t \geq X^s \boldsymbol{\lambda}$$

$$(1/\theta)\mathbf{y}_o^t \leq Y^s \boldsymbol{\lambda}$$

$$L \leq e\boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}.$$

In evaluating the within score $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^s)$, there are two schemes: one “inclusive” and the other “exclusive.” “Inclusive” for the input oriented model means that, when we evaluate $(\mathbf{x}_o, \mathbf{y}_o)^s$ with respect to the technology $(X, Y)^s$,

the DMU $(\mathbf{x}_o, \mathbf{y}_o)^s$ is always included in the evaluator $(X, Y)^s$, thus resulting in a score not greater than 1. The “Exclusive” scheme employs a method in which the DMU $(\mathbf{x}_o, \mathbf{y}_o)^s$ is removed from the evaluator group $(X, Y)^s$. This method of evaluation is equivalent to that of super-efficiency evaluation, and the score, if it exists, may be greater than 1. The intertemporal comparisons naturally utilize this “exclusive” scheme. So adoption of this scheme even in the within evaluations is not unnatural and promotes the discrimination power. Refer to Section 11.3.9 for further discussions on this subject.

In the BCC (VRS) model $[(L, U) = (1, 1)$: variable returns to scale], it may occur that the intertemporal LP (11.18) has no solution in its input and output orientations. In the case of the input-oriented model, (11.18) has no feasible solution if there exists i such that $y_{io}^t > \max_j \{y_{ij}^s\}$ whereas in the output-oriented case, (11.20) has no feasible solution if there exists i such that $x_{io}^t < \min_j \{x_{ij}^s\}$. Refer to Section 10.4 in Chapter 10 on this issue. In the IRS model $[(L, U) = (1, \infty)$: increasing returns to scale], it may occur that the output-oriented intertemporal LP has no solution, while the input-oriented case is always feasible. In case of DRS model $[(L, U) = (0, 1)$: decreasing returns to scale], it is possible that the input-oriented (11.18) has no solution, while the output-oriented model is always feasible. However, the CRS (CCR) model does not suffer from such trouble in its intertemporal measurement. Refer to Seiford and Zhu (1999)¹³ for infeasibility issues of radial super-efficiency models. One solution for avoiding this difficulty is to assign 1 to the score, since we have no means to evaluate the DMU within the evaluator group.

11.3.6 The Non-radial and Slacks-based MI

The radial approaches suffer from neglect of the slacks. In an effort to overcome this problem, several authors, e.g., Zhu (1996, 2001),¹⁴ Tone (2001, 2002)¹⁵ and Chen (2003)¹⁶ have developed non-radial measures of efficiency and super-efficiency. We develop here the non-radial, slacks-based MI. First, we introduce the input-oriented SBM (slack-based measure) and -SuperSBM (Tone (2001, 2002)). The SBM evaluates the efficiency of the examinee $(\mathbf{x}_o, \mathbf{y}_o)^s$ ($s = 1, 2$) with respect to the evaluator set $(X, Y)^t$ ($t = 1, 2$) with the following LP:

[SBM-I]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\lambda, s^-} 1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{io}^t \quad (11.21)$$

subject to

$$\mathbf{x}_o^t = X^s \boldsymbol{\lambda} + \mathbf{s}^-$$

$$\mathbf{y}_o^t \leq Y^s \boldsymbol{\lambda}$$

$$L \leq \mathbf{e} \boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}.$$

Or equivalently,

[SBM-I]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\boldsymbol{\xi}, \boldsymbol{\lambda}} 1 - \frac{1}{m} \sum_{i=1}^m \xi_i \quad (11.22)$$

subject to

$$(1 - \xi_i)x_{io}^t = \sum_{j=1}^n \lambda_j x_{ij}^s \quad (i = 1, \dots, m)$$

$$\mathbf{y}_o^t \leq Y^s \boldsymbol{\lambda}$$

$$L \leq e\boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\xi} \geq \mathbf{0}$$

where the vector $s^- \in R^m$ denotes input-slacks in (11.21). The equivalence between (11.21) and (11.22) can be shown as follows: Let $\xi_i = s_i^- / x_{io}^t$. Then equivalence follows in a straightforward manner. This model takes input slacks (surpluses) into account but not output slacks (shortfalls). Notice that, under the “inclusive” scheme (see the preceding section), [SBM-I] is always feasible for the case where $s = t$. However, under the “exclusive” scheme, we remove $(\mathbf{x}_o, \mathbf{y}_o)^t$ from the evaluator group $(X, Y)^t$ and hence [SBM-I] may have no feasible solution even for the case $s = t$. In this case, we solve the [SuperSBM-I] below:

[SuperSBM-I]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\boldsymbol{\xi}, \boldsymbol{\lambda}} 1 + \frac{1}{m} \sum_{i=1}^m \xi_i \quad (11.23)$$

subject to

$$(1 + \xi_i)x_{io}^t = \sum_{j=1}^n \lambda_j x_{ij}^s \quad (i = 1, \dots, m)$$

$$\mathbf{y}_o^t \leq Y^s \boldsymbol{\lambda}$$

$$L \leq e\boldsymbol{\lambda} \leq U$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\xi} \geq \mathbf{0}.$$

In this model, the score, if exists, satisfies $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) \geq 1$. In the output-oriented case, we solve the following programs.

[SBM-O]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\boldsymbol{\psi}, \boldsymbol{\lambda}} 1 / \left(1 + \frac{1}{m} \sum_{i=1}^m \psi_i \right) \quad (11.24)$$

subject to

$$\mathbf{x}_o^t \geq X^s \boldsymbol{\lambda}$$

$$(1 + \psi_i)y_{io}^t = \sum_{j=1}^n y_{ij}^s \lambda_j \quad (i = 1, \dots, q)$$

$$L \leq e\lambda \leq U$$

$$\lambda \geq 0, \psi \geq 0.$$

[SuperSBM-O]

$$\delta^s((x_o, y_o)^t) = \min_{\psi, \lambda} 1 / \left(1 - \frac{1}{q} \sum_{i=1}^q \psi_i \right) \quad (11.25)$$

subject to

$$x_o^t \geq X^s \lambda$$

$$(1 - \psi_i)y_{io}^t = \sum_{j=1}^n y_{ij}^s \lambda_j \quad (i = 1, \dots, q)$$

$$L \leq e\lambda \leq U$$

$$\lambda \geq 0, \psi \geq 0.$$

The output-oriented models take all output slacks (shortfalls) into account but no input slacks (excesses). The non-radial and slacks-based MI evaluates the four elements of MI: $\delta^1((x_o, y_o)^1)$, $\delta^2((x_o, y_o)^2)$, $\delta^1((x_o, y_o)^2)$ and $\delta^2((x_o, y_o)^1)$ by means of the programs [SBM-I] and [SuperSBM-I]. This is now explained below.

The inclusive scheme case

The within scores $\delta^1((x_o, y_o)^1)$ and $\delta^2((x_o, y_o)^2)$ are measured by solving [SBM-I] or [SBM-O]. They are not greater than 1. For the intertemporal scores $\delta^1((x_o, y_o)^2)$ and $\delta^2((x_o, y_o)^2)$, we solve the corresponding [SBM-I] or [SBM-O]. If they are feasible, the optimal value is the score that does not exceed 1. Otherwise, if they are infeasible, we apply [SuperSBM-I] or [SuperSBM-O] for measuring the score. In this case, the score, if exists, is always greater than 1.

The exclusive scheme case

To compute each component of MI, we apply [SBM-I] or [SBM-O] under the exclusive scheme. If they are feasible, the optimal value is the score that will never exceed 1. Otherwise, if they are infeasible, we solve [SuperSBM-I] or [SuperSBM-O] to compute the score. In this case, the score, if exists, is always greater than 1.

These models may suffer from the same infeasibility troubles as the radial ones.

11.3.7 The Non-radial and Non-oriented MI

The models in this category deal with input and output slacks. The [SBM] and [SuperSBM] models used for computing $\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t)$ are represented by the following fractional programs:

[SBM]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\phi, \psi, \lambda} \left(1 - \frac{1}{m} \sum_{i=1}^m \phi_i \right) / \left(1 + \frac{1}{q} \sum_{i=1}^q \psi_i \right) \quad (11.26)$$

subject to

$$(1 - \phi_i)x_{io}^t = \sum_{j=1}^n \lambda_j x_{ij}^s \quad (i = 1, \dots, m)$$

$$(1 + \psi_i)y_{io}^t = \sum_{j=1}^n \lambda_j y_{ij}^s \quad (i = 1, \dots, q)$$

$$L \leq e\lambda \leq U$$

$$\lambda \geq \mathbf{0}, \phi \geq \mathbf{0}, \psi \geq \mathbf{0}.$$

[SuperSBM]

$$\delta^s((\mathbf{x}_o, \mathbf{y}_o)^t) = \min_{\phi, \psi, \lambda} \left(1 + \frac{1}{m} \sum_{i=1}^m \phi_i \right) / \left(1 - \frac{1}{q} \sum_{i=1}^q \psi_i \right) \quad (11.27)$$

subject to

$$(1 + \phi_i)x_{io}^t = \sum_{j=1}^n \lambda_j x_{ij}^s \quad (i = 1, \dots, m)$$

$$(1 - \psi_i)y_{io}^t = \sum_{j=1}^n \lambda_j y_{ij}^s \quad (i = 1, \dots, q)$$

$$L \leq e\lambda \leq U$$

$$\lambda \geq \mathbf{0}, \phi \geq \mathbf{0}, \psi \geq \mathbf{0}.$$

These fractional programs can be transformed into LPs. See Section 10.3 in Chapter 10 or Tone (2002) for detailed discussions. This model under the exclusive scheme evaluates the four components of MI: $\delta^1((\mathbf{x}_o, \mathbf{y}_o)^1)$, $\delta^2((\mathbf{x}_o, \mathbf{y}_o)^2)$, $\delta^1((\mathbf{x}_o, \mathbf{y}_o)^2)$ and $\delta^2((\mathbf{x}_o, \mathbf{y}_o)^1)$ using [SBM], and, if the corresponding LP is found infeasible, we then apply [SuperSBM].

For this non-oriented model, [SuperSBM] is always feasible and has a finite minimum in any RTS environment, but under some mild conditions, i.e., for each output i ($= 1, \dots, q$) at least two DMUs must have positive values. This can be seen from the constraints in (11.27). See Tone (2002) for details or refer to Theorem 10.2 in Chapter 10.

11.3.8 Scale Efficiency Change

Changes in efficiency may also be associated with returns to scale efficiencies. We therefore treat this topic as follows: Let $\delta_C^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1})$ and $\delta_V^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1})$ denote the scores $\delta^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1})$ (11.13) that are obtained under CRS and VRS environments respectively. It holds, for any combination of t_1 and t_2 , that

$$\delta_C^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1}) \leq \delta_V^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1}). \quad (11.28)$$

We define the scale efficiency of $(\mathbf{x}_o, \mathbf{y}_o)^{t_1}$ relative to the technology $(X, Y)^{t_2}$ by

$$\sigma^{t_2}(\mathbf{x}_o, \mathbf{y}_o)^{t_1} = \delta_C^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1}) / \delta_V^{t_2}((\mathbf{x}_o, \mathbf{y}_o)^{t_1}). \quad (11.29)$$

From (11.28) it holds that $\sigma^{t_2}(\mathbf{x}_o, \mathbf{y}_o)^{t_1} \leq 1$ and, if $\sigma^{t_2}(\mathbf{x}_o, \mathbf{y}_o)^{t_1} = 1$, it means that $(\mathbf{x}_o, \mathbf{y}_o)^{t_1}$ is positioned in the most productive scale size (MPSS) region of the technology $(X, Y)^{t_2}$. See Banker (1984)¹⁷ and Banker *et al.* (1984).¹⁸ Hence, $\sigma^{t_2}(\mathbf{x}_o, \mathbf{y}_o)^{t_1}$ is an efficiency indicator showing how far $(\mathbf{x}_o, \mathbf{y}_o)^{t_1}$ deviates from the point of technically optimal scale of operation. Several authors (Färe *et al.* (1994),¹⁹ Lovell and Grifell-Tatje (1994),²⁰ Ray and Desli (1997),²¹ and Balk (2001)²²) have made attempts to enhance the Malmquist index by taking into account scale efficiency change effects. We here introduce Ray and Delsi (1997). See also Chapter 11 in Ray (2004).²³

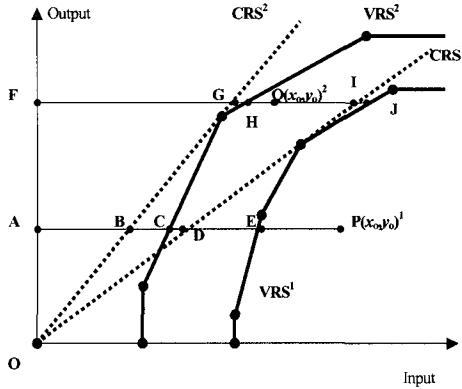
The Malmquist index in the CRS environment, MI_C , can be rewritten via the Malmquist index in VRS, MI_V , as follows:

$$MI_C = MI_V \times \left[\frac{\sigma^1(\mathbf{x}_o, \mathbf{y}_o)^2}{\sigma^1(\mathbf{x}_o, \mathbf{y}_o)^1} \times \frac{\sigma^2(\mathbf{x}_o, \mathbf{y}_o)^2}{\sigma^2(\mathbf{x}_o, \mathbf{y}_o)^1} \right]^{1/2}. \quad (11.30)$$

Ray and Delsi interpret the term in the square parenthesis as the geometric mean of scale efficiency changes of $(\mathbf{x}_o, \mathbf{y}_o)$ evaluated by the frontiers at the time periods 1 and 2. This can be illustrated in Figure 11.2 using a simple single-input and single-output case. The CRS frontier at time period 1 is the dotted line (ray) through the origin denoted by CRS¹, while the CRS frontier at 2 is the dotted line CRS². The VRS frontiers at 1(2) are the piecewise linear lines VRS¹ (VRS²). The points P($\mathbf{x}_o, \mathbf{y}_o$)¹ and Q($\mathbf{x}_o, \mathbf{y}_o$)² designate the DMU to be evaluated at 1 and 2 respectively. The numerator of the first fractional term in the parenthesis of (11.30) above is $(FI/FQ)/(FJ/FQ)=FI/FJ$, while the denominator is $(AE/AP)/(AC/AP)=AE/AC$. Hence the first fractional term turns out to be $(FI/FQ)/(AE/AC)$. This is the scale efficiency change from P($\mathbf{x}_o, \mathbf{y}_o$)¹ to Q($\mathbf{x}_o, \mathbf{y}_o$)² as evaluated by the $(X, Y)^1$ technology. Likewise, the second fractional term corresponds to the scale efficiency change evaluated by the $(X, Y)^2$ technology. Thus, their geometric mean represents the average scale efficiency change of $(\mathbf{x}_o, \mathbf{y}_o)$ from 1 to 2.

Ray and Delsi decompose MI_C in the following manner.

$$MI_C = \text{Catch-up(V)} \times \text{Frontier-shift(V)} \times \text{Scale Efficiency Change} \quad (11.31)$$

**Figure 11.2.** Scale Efficiency Change

where

$$\text{Catch-up}(V) = \frac{\delta_V^2((x_o, y_o)^2)}{\delta_V^1((x_o, y_o)^1)} \quad (11.32)$$

$$\text{Frontier-shift}(V) = \left[\frac{\delta_V^1((x_o, y_o)^1)}{\delta_V^2((x_o, y_o)^1)} \times \frac{\delta_V^1((x_o, y_o)^2)}{\delta_V^2((x_o, y_o)^2)} \right]^{1/2} \quad (11.33)$$

and

$$\text{Scale Efficiency Change} = \left[\frac{\sigma^1(x_o, y_o)^2}{\sigma^1(x_o, y_o)^1} \times \frac{\sigma^2(x_o, y_o)^2}{\sigma^2(x_o, y_o)^1} \right]^{1/2}. \quad (11.34)$$

11.3.9 Illustrative Examples for Model Comparisons

Using simple examples we here demonstrate and compare several models introduced in the preceding section.

1. Radial vs. Non-radial

Table 11.2 exhibits a small example with two DMUs, A and B, each having two inputs x_1 and x_2 , and a single output y , over two periods 1 and 2. At Period 1 DMU B has 1 unit of slacks in x_2 against A that is removed at Period 2. For brevity, we use the notation $\delta^s(A^t)$ for designating the efficiency score of DMU A at Period t relative to the production frontier at Period s . The scores measured by the input-oriented radial CRS and the non-radial CRS models under the inclusive scheme are displayed in Table 11.3. The radial

Table 11.2. Example 1

DMU	Period 1			Period 2		
	x_1	x_2	y	x_1	x_2	y
A	1	1	1	1	1	1
B	1	2	1	1	1	1

model records all scores as 1, whereas the non-radial (slacks-based) model records $\delta^1(B^1) = 0.75$ and $\delta^2(B^1) = 0.75$ reflecting the existence of slacks in x_2 at Period 1.

Table 11.3. Input-oriented Scores

	$\delta^1(A^1)$	$\delta^1(B^1)$	$\delta^1(A^2)$	$\delta^1(B^2)$	$\delta^2(A^1)$	$\delta^2(B^1)$	$\delta^2(A^2)$	$\delta^2(B^2)$
Radial	1	1	1	1	1	1	1	1
Non-radial	1	0.75	1	1	1	0.75	1	1

From Table 11.3, we obtain the catch-up (C), the frontier-shift (F) and the Malmquist index (MI) of each DMU from Period 1 to Period 2. These values are recorded in Table 11.4.

Table 11.4. Catch-up, Frontier-shift and Malmquist Index

	C(A)	C(B)	F(A)	F(B)	MI(A)	MI(B)
Radial	1	1	1	1	1	1
Non-radial	1	1.33	1	1	1	1.33

As can be seen from Table 11.2, DMU B has 1 unit of slack in input x_2 against A at Period 1, while the slacks are removed at Period 2. This effort is well captured in the non-radial model ($C(B)=1.33$), but the radial model neglects this removal, as well as the existence, of the slacks. In consequence, DMU B is seen as improving its productivity by 1.33 from Period 1 to Period 2 in the non-radial model, while the radial model reports no change in productivity, i.e., $MI(B)=1$.

2. Inclusive vs. Exclusive

In Section 11.3.5, we noted that, “inclusive” means that, when we evaluate $(x_o, y_o)^s$ with respect to the technology $(X, Y)^s$, the DMU $(x_o, y_o)^s$ is always included in the evaluator $(X, Y)^s$, thus resulting in the distance score being not greater than 1, while “exclusive” employs the scheme that the DMU $(x_o, y_o)^s$ is removed from the evaluator group $(X, Y)^s$. Here we compare the two schemes using an example. Table 11.5 exhibits a data set comprising three DMUs, A, B and C, with each one using a single input x to produce a single output y . Figure 11.3 is a diagrammatical representation of this data set, where the legend A¹ indicates DMU A at the period 1.

Table 11.5. Example 2

DMU	Period 1		Period 2	
	x	y	x	y
A	5	5	3	5
B	5	2	3	4
C	8	5	4	6

From the figure we can see that, under the CRS assumption, DMU A has an advantage over B and C in both periods. However, the degree of superiority is less in period 2 compared to period 1, implying that DMU A was “caught up” by its competitors.

First, we observe the catch-up effect of DMU A. The inclusive CRS models assign 1 to both $\delta^1(A^1)$ and $\delta^2(A^2)$, since A is on the efficient frontier of the CRS models at both periods. Hence, there is no catch-up effect, i.e., $C(A)=1$. On the other hand, the input-oriented exclusive non-radial CRS model evaluates the superiority A^1 over B^1 and C^1 as $\delta^1(A^1) = 1+3/5=1.67$ and the superiority of A^2 over B^2 and C^2 as $\delta^2(A^2) = 1+0.33/3 = 1.11$. Apparently, the superiority of A is diminished in period 2. The exclusive model thus renders $C(A) = 1.11/1.67=0.69$ which is less than 1. DMU A lost its superiority by about 40% at period 2. Second, both the inclusive and exclusive CRS models yield the intertemporal distance scores: $\delta^1(A^2) = 1.67$ and $\delta^2(A^1) = 0.6$. Using the frontier-shift formula (11.15), we have $F(A) = 1.67$ for the inclusive model and $F(A) = 2$ for the exclusive model. The difference is due to $\delta^1(A^1)$ and $\delta^2(A^2)$ for the two schemes. So we proceed further into the difference. We use the suffixes *incl* and *excl* for distinguishing the two schemes. Then, we have

$$\frac{F_{incl}(A)}{F_{excl}(A)} = \left[\left(\frac{\delta_{incl}^1(A^1)}{\delta_{excl}^1(A^1)} \right) \Bigg/ \left(\frac{\delta_{incl}^2(A^2)}{\delta_{excl}^2(A^2)} \right) \right]^{1/2}. \quad (11.35)$$

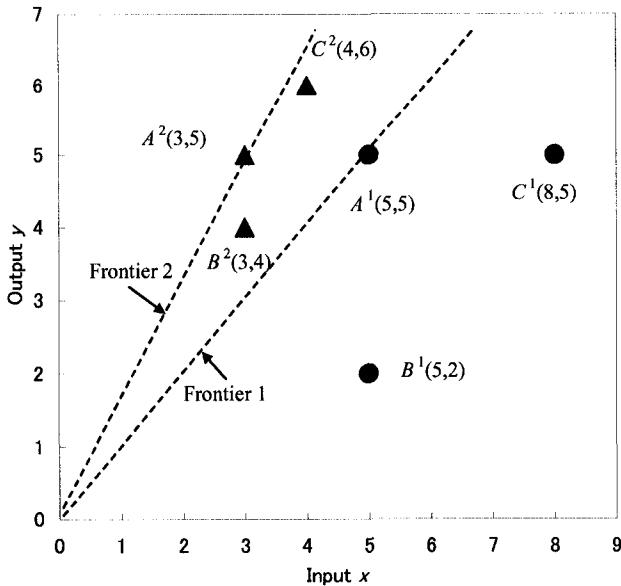


Figure 11.3. Three DMUs in Two Periods

Both terms in the parenthesis on the right of (11.35) are not greater than 1, since $\delta_{incl}^s(A^s) \leq \delta_{excl}^s(A^s)$ holds for $s = 1, 2$. If the first term is greater (less) than the second, then we have $F_{incl}(A) \geq (\leq) F_{excl}(A)$. In the case of our example, $1.67 = F_{incl}(A) < F_{excl}(A) = 2$.

Apart from this example, let us observe each component on the right of (11.35). If A^1 is positioned in the relative interior of $(X, Y)^1$, then we have $\delta_{incl}^1(A^1) = \delta_{excl}^1(A^1) < 1$. Hence no difference arises between the two. If A_1 is positioned on the efficient frontier of $(X, Y)^1$, then we have $\delta_{incl}^1(A^1) = 1$ and $\delta_{excl}^1(A^1) \geq 1$. In this case, $\delta_{excl}^1(A^1)$ is measured as the distance from A^1 to the set $(X, Y)^1$ excluding A^1 that may be called the distance from A^1 to the second best frontier of $(X, Y)^1$ around A^1 . Similarly, when A^2 is positioned on the efficient frontier of $(X, Y)^2$, then $\delta_{excl}^2(A^2) < 1$ indicates the distance from A^2 to the second best frontier of $(X, Y)^2$ around A^2 . Let us define $\alpha(\geq 1)$ and $\beta(\geq 1)$ by

$$\alpha = \frac{\delta_{excl}^1(A^1)}{\delta_{incl}^1(A^1)} \text{ and } \beta = \frac{\delta_{excl}^2(A^2)}{\delta_{incl}^2(A^2)}.$$

From (11.35), we have

$$F_{excl}(A) = F_{incl}(A) \times (\alpha/\beta)^{1/2}. \quad (11.36)$$

If $\alpha/\beta > 1(< 1)$, then the superiority of A is decreased (increased) against the second best frontier, and hence, the frontier-shift effect around A in-

creased (decreased) from Period 1 to Period 2 compared to those under the inclusive scheme evaluation.

The Malmquist indexes under these two schemes are respectively obtained as

$$MI_{incl}(A) = C_{incl}(A) \times F_{incl}(A) = 1 \times 1.67 = 1.67 \text{ and}$$

$$MI_{excl}(A) = C_{excl}(A) \times F_{excl}(A) = 0.69 \times 2 = 1.38.$$

As can be seen, in the exclusive scheme, the superiority of A over B and C is lost while its frontier-shift effect is increased. So, the Malmquist index value in the exclusive scheme is lower for DMU A compared to that in inclusive scheme, indicating a good catch-up by competitors in Period 2.

3. Oriented vs. Non-oriented

Table 11.6 exhibits three DMUs, A, B and C, each having two inputs (x_1 and x_2) and two outputs (y_1 and y_2) for two time periods. As can be seen, DMU A outperforms B and C in Period 1 since DMU B has one unit less of x_2 and DMU C one unit less of y_1 against A. These deficiencies are removed in Period 2, thus enabling these DMUs to reach an even status.

Table 11.6. Example 3

DMU	Period 1				Period 2			
	x_1	x_2	y_1	y_2	x_1	x_2	y_1	y_2
A	2	2	2	2	2	2	2	2
B	2	3	2	2	2	2	2	2
C	2	2	1	2	2	2	2	2

Using this data set, we compare four oriented models (input-oriented radial, input-oriented non-radial, output-oriented radial and output-oriented non-radial) and one non-oriented model (non-oriented non-radial) under the CRS environment. Table 11.7 displays the results. As expected, both the input-oriented and output-oriented radial models report no change as can be seen in the second and fourth columns of this table, where, as always, all DMUs are on the efficient frontiers from the radial standpoint. The input-oriented non-radial model assigns 1.2 to C(B) reflecting the removal of input slacks of DMU B in x_2 at Period 2. However, this model gives a score of 1 (no change) to C(C), F(C) and MI(C) ignoring the improvement of C in output y_1 at Period 2. The output-oriented non-radial model assigns a score of 1.5 to C(C), thus evaluating the removal of output shortfalls of C in y_1 at Period 2 but gives a score of 1 (no change) to C(B), F(B) and MI(B) neglecting B's effort for removing input slacks in x_2 . Finally, the non-oriented non-radial model takes account of all existing slacks, i.e., input excesses and output

shortfalls, and results in $MI(A)=0.89$, $MI(B)=1.2$ and $MI(C)=1.5$. As this example shows, the value of the Malmquist index is conditional upon the model selection, i.e., radial or non-radial, or, oriented or non-oriented. So due care is needed for deciding upon an appropriate model.

Table 11.7. Comparisons

	Input-oriented Radial	Input-oriented Non-radial	Output-oriented Radial	Output-oriented Non-radial	Non-oriented Non-radial
C(A)	1	0.8	1	0.75	0.8
C(B)	1	1.2	1	1	1.2
C(C)	1	1	1	1.5	1.5
F(A)	1	1.12	1	1.15	1.12
F(B)	1	1	1	1	1
F(C)	1	1	1	1	1
MI(A)	1	0.89	1	0.87	0.89
MI(B)	1	1.2	1	1	1.2
MI(C)	1	1	1	1.5	1.5

4. Infeasibility LP issues

As noted in Section 11.3.5, the oriented models in the VRS environment may suffer from the problem of infeasibility as in the measurement of the super-efficiency. However, the non-oriented VRS model is free from this trouble. We now demonstrate this fact with the help of the small data set reported in Table 11.8.

Table 11.8. Example 4

DMU	Period 1		Period 2	
	x	y	x	y
A	3	1	1	3
B	4	2	2	4

In the oriented VRS model, we must solve the intertemporal scores $\delta^2(A^1)$, $\delta^2(B^1)$, $\delta^1(A^2)$ and $\delta^1(B^2)$, defined by (11.18), (11.20), (11.23) or (11.25). We encounter difficulties in solving LPs for this example. Specifically, in the input-oriented case, (11.18) and (11.23) have no feasible solution for $\delta^1(A^2)$ and $\delta^1(B^2)$, since $y_A^2 > \max\{y_A^1, y_B^1\}$ and $y_B^2 > \max\{y_A^1, y_B^1\}$. In the output-oriented case, (11.20) and (11.25) have no feasible solution for $\delta^1(A^2)$ and

$\delta^1(B^2)$, since $x_A^2 < \min\{x_A^1, x_B^1\}$ and $x_B^2 < \min\{x_A^1, x_B^1\}$. Thus, in both cases, we cannot evaluate the frontier-shift effects and hence the Malmquist indexes are not available for all DMUs A and B. The non-oriented VRS model (11.27), however, can always measure these scores, as is shown in Table 11.9.

Table 11.9. Results by the Non-oriented Non-radial Model

	$\delta^1(.^1)$	$\delta^1(.^2)$	$\delta^2(.^1)$	$\delta^2(.^2)$	Catch-up	Frontier-shift	Malmquist
A	1.33	6	0.11	2	1.5	6	9
B	2	4	0.17	1.33	0.67	6	4

11.3.10 Concluding Remarks

In this chapter we have discussed the Malmquist productivity index, and its measurements along with scale efficiency change and related subjects. All of the approaches we have covered in our opening paragraph utilize the ratio forms required in (11.16) with the exception of Thrall (2000) who provides an additive model. However, Thrall's formulation might best be regarded as an alternative to the MI and has therefore not been covered in this chapter. We conclude this chapter by pointing to several crucial issues involved in applying the Malmquist index model.

1. Selection of Types of Returns-to-Scale

We need to exercise due care in selecting an appropriate RTS type. This identification problem is common to all DEA applications, since DEA is a data-oriented and non-parametric method. If the data set includes numeric values with a large difference in magnitude, e.g., comparing big companies with small ones, the VRS model may be a choice. However, if the data set consists of normalized numbers, e.g., per capita acre and hour, the CRS model might be an appropriate candidate.

2. Radial or Non-radial

The radial Malmquist index is based upon the radial DEA scores and hence remaining non-zero slacks are not counted in the radial scores. If slacks are not "freely disposable," the radial Malmquist index cannot fully characterize the productivity change. In contrast, the input (output) oriented non-radial Malmquist index takes into account the input (output) slacks. This results in a smaller objective function value than in the radial model both in the Within and Intertemporal scores. Chen (2003) applied the non-radial Malmquist model to measure the productivity change of three major Chinese industries: Textiles, Chemicals and Metallurgicals. The results indicate that it is important to consider the possible non-zero slacks in measuring the

productivity change. The radial model could not explain and hence could not fully capture what the data say because of the large amount of non-zero slacks.

3. Infeasibility and its Solution

As pointed out in Sections 11.3.5 and 11.3.6, the oriented VRS, IRS and DRS models suffer from the problem of infeasible solution in the Intertemporal score (super-efficiency) estimation. This is not a rare case, since the Malmquist model deals with data sets over time that can display large variations in magnitude and hence the super-efficiency model suffers from many infeasible solutions. However, the Non-radial and Non-oriented Model is free from such problems.

4. Empirical Studies

The Malmquist indexes have been used in a variety of studies. Färe *et al.* (1998a)²⁴ surveyed empirical applications comprising various sectors such as agriculture, airlines, banking, electric utilities, insurance companies, and public sectors. However, most of these studies utilized the oriented-radial models. Thus, they suffer from the neglect of slacks and infeasibility. Generally speaking, however, empirical studies are at a nascent stage as of now.

11.4 SUMMARY OF CHAPTER 11

In this chapter, we introduced two methods for measuring efficiency change over time: “window analysis” and “Malmquist index” using non-parametric DEA models.

11.5 NOTES AND SELECTED BIBLIOGRAPHY

The name and the basic concept of “window analysis” is due to G. Klopp (1985).²⁵

The most important paper on the Malmquist index is R. Färe, S. Grosskopf, S. Norris and Z. Zhang (1994), “Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries,” *The American Economic Review* 84, pp.66-83, in which during the period covered they showed that U.S. was first in changing the frontier and Japan was first in catching up.

The non-radial Malmquist indices were proposed by Tone (2004).²⁶

11.6 RELATED DEA-SOLVER MODELS FOR CHAPTER 11

Window-I(O)-C(V) These codes execute Window Analysis in Input (Output) orientation under constant (CRS) or variable (VRS) returns-to-scale assumptions. See the sample data format in Section B.5.9 and explanation on results in Section B.7 of Appendix B.

Malmquist-Radial-I(O)-C(V,GRS) These codes solve Radial and Oriented MI under the Constant, Variable or General returns-to-scale assumptions. Data format is the same with the Window Analysis. See Section B.5.9.

Malmquist-I(O)-C(V,GRS) These codes solve Non-radial and Oriented MI under the Constant, Variable or General returns-to-scale assumptions. Data format is the same with the Window Analysis. See Section B.5.9.

Malmquist-C(V,GRS) These codes solve Non-radial and Non-oriented MI under the Constant, Variable or General returns-to-scale assumptions. Data format is the same with the Window Analysis. See Section B.5.9.

Notes

1. This work was subsequently incorporated in his Ph. D. thesis: G. Klopp (1985), "The Analysis of the Efficiency of Production System with Multiple Inputs and Outputs" (Chicago: University of Illinois at Chicago, Industrial and Systems Engineering College). Also available from University Microfilms, Inc., 300 North Zeeb Road, Ann Arbor, Mich. 48106. See also A. Charnes, T. Clark, W.W. Cooper and B. Golany (1985) "A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Force," *Annals of Operations Research* 2, pp.95-112. See also W.F. Bowlin (1984) "A Data Envelopment Analysis Approach to Performance Evaluation in Not-for Profit Entities with an Illustrative Application to the U.S. Air Force," Ph. D. Thesis (Austin, Texas: The University of Texas, Graduate School of Business). Also available from University Microfilms, Inc.
2. S. Malmquist (1953), "Index Numbers and Indifference Surfaces," *Trabajos de Estadística* 4, pp.209-242.
3. D.W. Caves, L.R. Christensen and W.E. Diewert (1982), "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica* 50, pp.1393-1414.
4. R. Färe and S. Grosskopf (1992), "Malmquist Indexes and Fisher Ideal Indexes," *The Economic Journal* 102, pp.158-160.
5. R. Färe, S. Grosskopf, B. Lindgren and P. Roos (1989), (1994), "Productivity Change in Swedish Hospitals: A Malmquist Output Index Approach," in Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (eds.) *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston, Kluwer Academic Publishers.
6. R. Färe, S. Grosskopf and R. Russell, (eds.), (1998b), *Index Numbers: Essays in Honour of Sten Malmquist*, Kluwer Academic Publishers.
7. R.M. Thrall (2000), "Measures in DEA with an Application to the Malmquist Index," *Journal of Productivity Analysis* 13, pp.125-137.
8. As determined from YATS (Youth Attitude Tracking Survey) which obtains this information from periodic surveys conducted for the Army.
9. T. Sueyoshi (1992), "Comparisons and Analyses of Managerial Efficiency and Returns to Scale of Telecommunication Enterprises by using DEA/WINDOW," (in Japanese) *Communications of the Operations Research Society of Japan* 37, pp.210-219.
10. D.B. Sun (1988), "Evaluation of Managerial Performance in Large Commercial Banks by Data Envelopment Analysis," Ph.D. Thesis (Austin, Texas: The University of Texas, Graduate School of Business). Also available from University Microfilms, Inc. See Footnote 1.
11. A. Charnes and W.W. Cooper (1991), "DEA Usages and Interpretations" reproduced in *Proceedings of International Federation of Operational Research Societies 12th Triennial Conference* in Athens, Greece, 1990.
12. P. Andersen and N.C. Petersen (1993), "A Procedure for Ranking Efficient Units in Data Envelopment Analysis," *Management Science* 39, pp.1261-1264.
13. L.M. Seiford and J. Zhu (1999), "Infeasibility of Super-efficiency Data Envelopment Analysis," *INFOR* 37, pp.174-187.
14. J. Zhu (1996), "Data Envelopment Analysis with Preference Structure," *Journal of the Operational Research Society* 47, pp.136-150.

15. K. Tone (2001), "A Slacks-based Measure of Efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 130, pp.498-509. K. Tone (2002), "A Slacks-based Measure of Super-efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 143, pp.32-41.
16. Y. Chen (2003), "Non-radial Malmquist Productivity Index with an Illustrative Application to Chinese Major Industries," *International Journal of Production Economics* 83, pp.27-35.
17. R.D. Banker, (1984), "Estimating the Most Productive Scale Size using Data Envelopment Analysis," *European Journal of Operational Research* 17, pp.35-44.
18. R.D. Banker, A. Charnes and W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30, pp.1078-1092.
19. R. Färe, S. Grosskopf, M. Norris and Z. Zhang (1994), "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *The American Economic Review* 84, pp.66-83.
20. C.A.K. Lovell and E. Grifell-Tatje (1994), "A Generalized Malmquist Productivity Index," Paper presented at the Georgia Productivity Workshop at Athens, GA, October 1994.
21. S.C. Ray and E. Delsi (1997), "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment," *The American Economic Review* 87, pp.1033-1039.
22. B.M. Balk (2001), "Scale Efficiency and Productivity Change," *Journal of Productivity Analysis* 15, pp.159-183.
23. S.C. Ray (2004) *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, Cambridge University Press.
24. R. Färe, S. Grosskopf and P. Roos (1998a), "Malmquist Productivity Indexes: A Survey of Theory and Practice," in *Index Numbers: Essays in Honour of Sten Malmquist*, Kluwer Academic Publishers, pp.127-190.
25. See Note 1.
26. K. Tone (2004) "Malmquist Productivity Index: Efficiency Change Over Time," Chapter 8 in *Handbook on Data Envelopment Analysis* edited by W. W. Cooper, L. M. Seiford and J. Zhu (Norwell Mass., Kluwer Academic Publishers).

12 SCALE ELASTICITY AND CONGESTION

12.1 INTRODUCTION

Most of the business units (or decision making units (DMUs)) today face an increasingly limited supply of resources and persistent competition. This has led to a significant emphasis on the efficient utilization and allocation of on-hand resources by building larger operating units to achieve the possible advantages of ‘scale economies.’ From a policy point of view, the estimation of scale elasticity (returns to scale) parameter is of particular importance concerning whether there is any scope for increased productivity by expanding/contracting, whether minimum efficient scales will allow competitive markets to be established, and if the existing size distribution of firms is consistent with a competitive market outcome. We have already discussed this subject qualitatively in Chapter 4 with an emphasis on the most productive scale size (MPSS). Here, a quantitative analysis on this subject will be introduced.

There are many resources which affect the performance of a unit when there is ‘overuse’ of such resources. When firms use these resources they only take into account their own benefits and costs of such usage, but largely ignore the congestion, or exclusion costs that they impose on others. This is referred to as “congestion externality” in the economics literature. When congestion is present, it effectively shrinks business market areas and reduces the scale economies. So there is a need to estimate the returns to scale parameter in the

presence of congestion, and to examine it with more prudence for the firm's financial viability and success.

This chapter unfolds as follows. Section 2 discusses the scale elasticity or the degree of scale economies (DSE). Then we turn our attention to the problem of scale elasticity in the presence of congestion, and introduce a method for identifying DSE in Section 3. Sections 4 introduces illustrative examples. This topic is covered mainly from Tone and Sahoo (2004).¹

12.2 SCALE ELASTICITY IN PRODUCTION

Throughout this paper, we deal with n DMUs, each having m inputs for producing s outputs. For each DMU _{o} ($o = 1, \dots, n$), we denote respectively the input/output vectors by $x_o \in R^m$ and $y_o \in R^s$. The input/output matrices are defined by $X = (x_1, \dots, x_n) \in R^{m \times n}$ and $Y = (y_1, \dots, y_n) \in R^{s \times n}$. We assume that $X > O$ and $Y > O$.

The returns to scale (RTS) or scale elasticity in production (ρ) or degree of scale economies (DSE) is defined as the ratio of marginal product (MP) to average product (AP). In a single input/output case, if the output y is produced by the input x , we define the scale elasticity ρ by

$$\rho = \text{MP/AP} = \frac{dy}{dx} / \frac{y}{x}. \quad (12.1)$$

(Refer to the beginning discussions of Section 4.3.)

For a neoclassical 'S-shaped production function' (or *Regular Ultra Passum Law* (RUPL) in the words of Frisch, 1965²), ρ can take on values ranging from 'greater than one' to 'less than one,' and even negative values when production decreases with usage of inputs. RTS is said to be *increasing*, *constant* and *decreasing* if $\rho > 1$, $\rho = 1$ and $\rho < 1$ respectively.

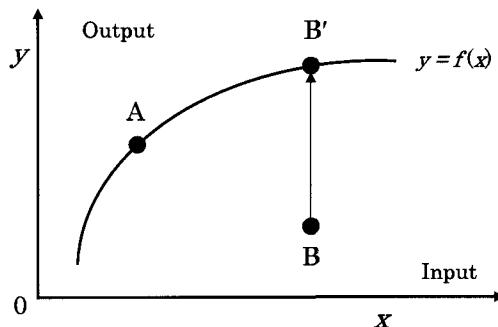


Figure 12.1. Scale Elasticity

Figure 12.1 exhibits such a sample curve $y = f(x)$ to demonstrate scale elasticity in production. Scale elasticity is well-defined at a point on the efficient

portion of the input-output correspondence, e.g., the point A. For an inefficient DMU operating on point such as B, ρ is defined on a point such as its upward projected point B' which is on the efficient frontier.

In the case of multiple input-output environment, although Baumol *et al.* (1988)³ discussed DSE in terms of cost and output, we utilize the same terminology in our case. The reason is that any differentiable cost function, whatever the number of outputs involved, and whether or not it is derived from a homogeneous production process, has a local degree of homogeneity (or, equivalently DSE) that is reciprocal of RTS parameter of a production process (p.56). We deal with the production possibility set P_{BCC} related to the multiple input-output correspondence defined by

$$P_{BCC} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y} \leq Y\boldsymbol{\lambda}, \mathbf{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}, \quad (12.2)$$

where \mathbf{e} is a row vector with all its elements being equal to one.

Analogous to the single input-output case, we measure the scale elasticity of DMUs positioned on the efficient portion of P_{BCC} . Such projection can be realized by solving the following output-oriented BCC model (Banker *et al.*, 1984, 1992)⁴ for each DMU $(\mathbf{x}_o, \mathbf{y}_o)$ ($o = 1, \dots, n$).

$$\begin{aligned} [\text{BCC-O}] \quad & \theta_{BCC}^* = \max \theta_{BCC} \\ \text{subject to} \quad & \mathbf{x}_o = X\boldsymbol{\lambda} + \mathbf{s}^- \\ & \theta_{BCC}\mathbf{y}_o = Y\boldsymbol{\lambda} - \mathbf{s}^+ \\ & \mathbf{e}\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}. \end{aligned} \quad (12.3)$$

To solve [BCC-O], we employ a two-stage process: first we obtain θ_{BCC}^* by solving the program [BCC-O], and then we maximize the sum of input slacks s_i^- and output slacks s_r^+ while keeping the objective function value at θ_{BCC}^* . Refer to Chapter 3 for the detailed explanation of this two-stage evaluation process.

Let an optimal solution of [BCC-O] be $(\theta_{BCC}^*, \boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$. Then the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is called “strongly efficient” if and only if $\theta_{BCC}^* = 1$, $\mathbf{s}^{-*} = \mathbf{0}$ and $\mathbf{s}^{+*} = \mathbf{0}$ hold for every optimal solution for [BCC-O]. Otherwise, the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ can be brought into the strongly efficient status by the BCC-O projection as defined by

$$\mathbf{x}_o^* \leftarrow \mathbf{x}_o - \mathbf{s}^{-*} \quad (12.4)$$

$$\mathbf{y}_o^* \leftarrow \theta_{BCC}^*\mathbf{y}_o + \mathbf{s}^{+*}. \quad (12.5)$$

Hereafter, we assume that this operation has been performed so DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is strongly efficient.

The dual program to the [BCC-O] model, using the dual variables $\mathbf{v} \in R^m$, $\mathbf{u} \in R^s$ and $w \in R$, is described as follows:

$$[\text{Dual}] \quad \min \mathbf{v}\mathbf{x}_o - w \quad (12.6)$$

$$\text{subject to } -\mathbf{v}X + \mathbf{u}Y + \mathbf{e}w \leq \mathbf{0} \quad (12.7)$$

$$\mathbf{u}\mathbf{y}_o = 1 \quad (12.8)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, w : \text{free in sign.} \quad (12.9)$$

Since we assume that the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is (strongly) efficient, it follows from the duality theorem of linear programming that there exists an optimal solution $(\mathbf{v}^*, \mathbf{u}^*, w^*)$ for [Dual] such that

$$\mathbf{v}^* \mathbf{x}_o - w^* = 1. \quad (12.10)$$

From (12.7), (12.8) and (12.10), it holds, for the optimal $(\mathbf{v}^*, \mathbf{u}^*, w^*)$, that

$$-\mathbf{v}^* \mathbf{x}_j + \mathbf{u}^* \mathbf{y}_j + w^* \leq 0 \quad (j = 1, \dots, n) \quad (12.11)$$

$$-\mathbf{v}^* \mathbf{x}_o + \mathbf{u}^* \mathbf{y}_o + w^* = 0 \quad (12.12)$$

$$\mathbf{v}^* \geq \mathbf{0}, \mathbf{u}^* \geq \mathbf{0}. \quad (12.13)$$

This means that the hyperplane $-\mathbf{v}^* \mathbf{x} + \mathbf{u}^* \mathbf{y} + w^* = 0$ is a supporting hyperplane to the production possibility set P_{BCC} at $(\mathbf{x}_o, \mathbf{y}_o)$.

We now define respectively the virtual input (ξ) and virtual output (η) associated with $(\mathbf{x}, \mathbf{y}) \in P_{BCC}$ by

$$\xi = \mathbf{v}^* \mathbf{x} \text{ and } \eta = \mathbf{u}^* \mathbf{y}. \quad (12.14)$$

Thus, the hyperplane can be expressed as

$$-\xi + \eta + w^* = 0. \quad (12.15)$$

From this equation, we derive marginal product (MP) as

$$\text{MP} = \frac{d\eta}{d\xi} = 1, \quad (12.16)$$

and average product (AP) as

$$\text{AP} = \frac{\eta}{\xi} = \frac{\eta}{\eta + w^*}. \quad (12.17)$$

At $(\mathbf{x}_o, \mathbf{y}_o)$, we have $\eta = 1$ from (12.8) and (12.14), and hence the scale elasticity (ρ) or DSE at $(\mathbf{x}_o, \mathbf{y}_o)$ is expressed by

$$\rho = 1 + w^*. \quad (12.18)$$

On many occasions, the optimal w^* is not uniquely determined, i.e., there exist multiple optima. In such cases, we can find the upper (lower) bound \bar{w} (\underline{w}) of w by solving the following linear program in \mathbf{v} , \mathbf{u} and w .

$$[\text{Upper(Lower)}] \quad \bar{w}(\underline{w}) = \max(\min)w \quad (12.19)$$

$$\text{subject to } -\mathbf{v}X + \mathbf{u}Y + \mathbf{e}w \leq \mathbf{0} \quad (12.20)$$

$$-\mathbf{v}\mathbf{x}_o + \mathbf{u}\mathbf{y}_o + w = 0 \quad (12.21)$$

$$\mathbf{u}\mathbf{y}_o = 1 \quad (12.22)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, w : \text{free in sign.} \quad (12.23)$$

The upper (lower) scale elasticity in production $\bar{\rho}$ ($\underline{\rho}$) is calculated by

$$\bar{\rho} = 1 + \bar{w} \text{ and } \underline{\rho} = 1 + \underline{w}. \quad (12.24)$$

Regarding the scale elasticity in production, we have a units invariance theorem as follows:

Theorem 12.1 *The scale elasticity in production (ρ) as defined by (12.18), and its upper (lower) bound $\bar{\rho}$ ($\underline{\rho}$) are all units invariant.*

See Problem 12.1 for a proof.

12.3 CONGESTION

So far we have dealt with situations where input slacks (excesses) are considered free. The set P_{BCC} allows an (unbounded) input $x \geq X\lambda$ for producing an output $y = Y\lambda$. Under this assumption, the scale elasticity ρ is nonnegative, since we have

$$\rho = 1 + w^* = v^* x_o \geq 0.$$

However, there are some cases in which an increase in one or more inputs causes the worsening of one or more outputs. A typical example is the case of mining where too many miners in an underground mine may lead to "congestion" with reference to output. Figure 12.2 exhibits such a phenomenon. An increase in input x results in a decrease in output y as is shown in the case of points such as F and G.

In order to potentially deal with such situations, we need to modify our production possibility set as follows:

$$P_{convex} = \{(x, y) | x = X\lambda, y \leq Y\lambda, e\lambda = 1, \lambda \geq 0\}. \quad (12.25)$$

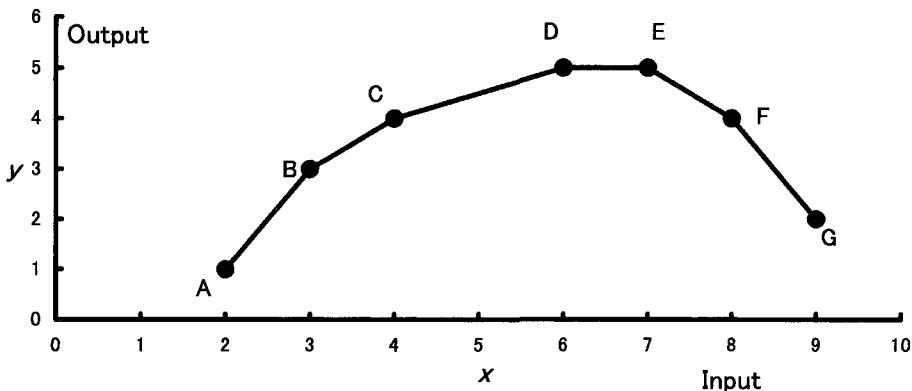


Figure 12.2. Congestion

In this section, we first discuss the scale elasticity issue with respect to this new production possibility set P_{convex} , and then demonstrate that “(strong) congestion” is recognized by the status by having negative production elasticity ($\rho < 0$). The degree of congestion can be measured by ρ , in Section 12.3.1. However, this condition ($\rho < 0$) is, in some cases, too restrictive to decide upon the status of congestion. We, therefore, introduce the status of “weak congestion” in Section 12.3.2. Then, a summary of the degree of scale economies and the status of congestion is presented in Section 12.3.3.

12.3.1 Strong Congestion

Similar to the P_{BCC} case, we assume here that the concerned DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is on the strongly efficient frontier of P_{convex} . This means that the following model has a solution ($\phi^* = 1, \mathbf{q}^{+*} = \mathbf{0}$) for every optimum.

$$\begin{aligned} [\text{Congestion-strong}] \quad & \max \phi && (12.26) \\ \text{subject to} \quad & \mathbf{x}_o = X\boldsymbol{\lambda} \\ & \phi\mathbf{y}_o = Y\boldsymbol{\lambda} - \mathbf{q}^+ \\ & e\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{q}^+ \geq \mathbf{0}. \end{aligned}$$

If the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is not efficient, we project it onto the efficient frontier of P_{convex} by the following formulas:

$$\mathbf{x}_o^* \leftarrow \mathbf{x}_o \text{ unchanged} \quad (12.27)$$

$$\mathbf{y}_o^* \leftarrow \phi^*\mathbf{y}_o + \mathbf{q}^{+*}. \quad (12.28)$$

The input-output vector $(\mathbf{x}_o^*, \mathbf{y}_o^*)$ is now strongly P_{convex} -efficient.

The dual program of [Congestion-strong] is described as follows:

$$[\text{Dual-Congestion}] \quad \min v\mathbf{x}_o - w \quad (12.29)$$

$$\text{subject to} \quad -vX + uY + ew \leq \mathbf{0} \quad (12.30)$$

$$u\mathbf{y}_o = 1 \quad (12.31)$$

$$u \geq \mathbf{0}, v, w : \text{free in sign.} \quad (12.32)$$

We notice that, in this dual program, v is free in sign, whereas it was constrained to be nonnegative in [Dual]. See (12.9). Let an optimal solution vector of [Dual-Congestion] be (v^*, u^*, w^*) . Then, we can demonstrate in a vein similar to P_{BCC} that the hyperplane $-v^*\mathbf{x} + u^*\mathbf{y} + w^* = 0$ is a supporting hyperplane to P_{convex} at $(\mathbf{x}_o, \mathbf{y}_o)$. The production elasticity can then be obtained by

$$\rho = 1 + w^*. \quad (12.33)$$

The multiple optima case can be dealt with by the following program:

$$[\text{Upper-w}] \quad \bar{w} = \max w \quad (12.34)$$

$$\text{subject to } -\mathbf{v}X + \mathbf{u}Y + \mathbf{e}w \leq \mathbf{0} \quad (12.35)$$

$$-\mathbf{v}\mathbf{x}_o + \mathbf{u}\mathbf{y}_o + w = 0 \quad (12.36)$$

$$\mathbf{u}\mathbf{y}_o = 1 \quad (12.37)$$

$$\mathbf{u} \geq \mathbf{0}, \mathbf{v}, w : \text{free in sign}. \quad (12.38)$$

Here we have an interest in the upper bound values of w^* , since this value is closely related with “congestion.” Thus, in a manner similar to (12.19)-(12.23), the upper bound of scale elasticity is calculated by

$$\bar{\rho} = 1 + \bar{w}. \quad (12.39)$$

The dual program to [Upper-w] can be described as follows:

$$[\text{Dual-Upper-w}] \quad \min \theta_1 - \theta_2 \quad (12.40)$$

$$\text{subject to } X\boldsymbol{\lambda} = \theta_1 \mathbf{x}_o \quad (12.41)$$

$$Y\boldsymbol{\lambda} \geq \theta_2 \mathbf{y}_o \quad (12.42)$$

$$\mathbf{e}\boldsymbol{\lambda} = 1 + \theta_1 \quad (12.43)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \theta_1, \theta_2 : \text{free in sign}. \quad (12.44)$$

Let an optimal solution vector of [Dual-Upper-w] be $(\theta_1^*, \theta_2^*, \boldsymbol{\lambda}^*)$. Then, the upper bound of scale elasticity $\bar{\rho}$ satisfies the following equality:

$$\bar{\rho} = 1 + \theta_1^* - \theta_2^*. \quad (12.45)$$

Now, suppose that $\bar{\rho} < 0$, i.e., $1 + \theta_1^* - \theta_2^* < 0$. In this case, from the constraints in [Dual-Upper-w] and the assumptions $X > O$ and $Y > O$, we have:

$$\theta_2^* > 1 + \theta_1^* > \theta_1^* > 0 \text{ and } \boldsymbol{\lambda}^* \neq \mathbf{0}. \quad (12.46)$$

Let us define $\hat{\boldsymbol{\lambda}}$, $\hat{\mathbf{x}}_o$ and $\hat{\mathbf{y}}_o$ respectively as

$$\hat{\boldsymbol{\lambda}} = \frac{1}{1 + \theta_1^*} \boldsymbol{\lambda}^*, \hat{\mathbf{x}}_o = X\hat{\boldsymbol{\lambda}} \text{ and } \hat{\mathbf{y}}_o = Y\hat{\boldsymbol{\lambda}}. \quad (12.47)$$

Then, we have

$$\mathbf{e}\hat{\boldsymbol{\lambda}} = 1 \quad (12.48)$$

$$X\hat{\boldsymbol{\lambda}} = \hat{\mathbf{x}}_o = \frac{\theta_1^*}{1 + \theta_1^*} \mathbf{x}_o < \mathbf{x}_o \quad (12.49)$$

$$Y\hat{\boldsymbol{\lambda}} = \hat{\mathbf{y}}_o \geq \frac{\theta_2^*}{1 + \theta_1^*} \mathbf{y}_o > \mathbf{y}_o. \quad (12.50)$$

Thus, there exists $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o) \in P_{\text{convex}}$ such that $\hat{\mathbf{x}}_o < \mathbf{x}_o$ and $\hat{\mathbf{y}}_o > \mathbf{y}_o$. This means that the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is in the region of “congestion” where there exists a different activity $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ that uses less inputs to produce more outputs.

Conversely, suppose that there exists $(\tilde{\mathbf{x}}_o, \tilde{\mathbf{y}}_o) \in P_{\text{convex}}$ such that

$$\tilde{\mathbf{x}}_o = \alpha \mathbf{x}_o \text{ (with } 0 < \alpha < 1\text{)} \text{ and } \tilde{\mathbf{y}}_o \geq \beta \mathbf{y}_o \text{ (with } \beta > 1\text{)}. \quad (12.51)$$

Then, we can prove that $(\mathbf{x}_o, \mathbf{y}_o)$ has a negative upper bound of scale elasticity, as follows:

Since adding the activity $(\tilde{\mathbf{x}}_o, \tilde{\mathbf{y}}_o)$ to the data set (X, Y) has no effect on the curvature of the set P_{convex} , we enlarge the data set (X, Y) to (\bar{X}, \bar{Y}) by adding $(\tilde{\mathbf{x}}_o)$ and $(\tilde{\mathbf{y}}_o)$ to the first column of X and Y , respectively. Thus, we have

$$\bar{X} = (\tilde{\mathbf{x}}_o, \mathbf{x}_1, \dots, \mathbf{x}_n) \in R^{m \times (n+1)} \text{ and } \bar{Y} = (\tilde{\mathbf{y}}_o, \mathbf{y}_1, \dots, \mathbf{y}_n) \in R^{s \times (n+1)}.$$

Let us define the corresponding $\bar{\lambda} \in R^{n+1}$ by

$$\bar{\lambda}^T = (\lambda_1, 0, 0, \dots, 0).$$

Then, we have

$$\bar{X}\bar{\lambda} = \lambda_1 \tilde{\mathbf{x}}_o = \lambda_1 \alpha \mathbf{x}_o \quad (12.52)$$

$$\bar{Y}\bar{\lambda} = \lambda_1 \tilde{\mathbf{y}}_o \geq \lambda_1 \beta \mathbf{y}_o \quad (12.53)$$

$$\mathbf{e}\bar{\lambda} = \lambda_1. \quad (12.54)$$

Let us define λ_1 , θ_1 and θ_2 respectively by

$$\lambda_1 = \frac{1}{1 - \alpha} \quad (12.55)$$

$$\theta_1 = \alpha \lambda_1 = \frac{\alpha}{1 - \alpha} \quad (12.56)$$

$$\theta_2 = \beta \lambda_1 = \frac{\beta}{1 - \alpha}. \quad (12.57)$$

Then, we have

$$\bar{X}\bar{\lambda} = \theta_1 \mathbf{x}_o \quad (12.58)$$

$$\bar{Y}\bar{\lambda} \geq \theta_2 \mathbf{y}_o \quad (12.59)$$

$$\mathbf{e}\bar{\lambda} = 1 + \theta_1 \quad (12.60)$$

$$\bar{\lambda} \geq 0. \quad (12.61)$$

Thus, this solution vector $(\theta_1, \theta_2, \bar{\lambda})$ is feasible in [Dual-Upper-w]. For this feasible solution we now have

$$1 + \theta_1 - \theta_2 = \frac{1 - \beta}{1 - \alpha} < 0, \text{ since } \alpha < 1 \text{ and } \beta > 1 \text{ by assumption.} \quad (12.62)$$

Hence, it is proved that the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ has a negative upper bound of scale elasticity.

To conclude this section, we have the following definition and theorem.

Definition 12.1 (Strong Congestion)

A DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is strongly congested if there exists an activity $(\tilde{\mathbf{x}}_o, \tilde{\mathbf{y}}_o) \in P_{convex}$ such that $\tilde{\mathbf{x}}_o = \alpha \mathbf{x}_o$ (with $0 < \alpha < 1$) and $\tilde{\mathbf{y}}_o \geq \beta \mathbf{y}_o$ (with $\beta > 1$).

Theorem 12.2 (Tone and Sahoo (2004)) A DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is in the status of strong congestion, if and only if the upper scale elasticity ($\bar{\rho}$) as measured by the scheme [Upper-w] is negative.

We notice that this definition of strong congestion requires existence of a DMU $(\tilde{\mathbf{x}}_o, \tilde{\mathbf{y}}_o) \in P_{convex}$ with a proportionally reduced input vector $\tilde{\mathbf{x}}_o$ ($= \alpha \mathbf{x}_o$, $0 < \alpha < 1$) and an enlarged output vector $\tilde{\mathbf{y}}_o$ ($\geq \beta \mathbf{y}_o$, $\beta > 1$).

12.3.2 Weak Congestion

The common understanding of congestion is that an increase (decrease) in one or more inputs causes a decrease (increase) in one or more outputs (Cooper *et al.*, 2001).⁵ From this point of view, the above definition of strong congestion is too restrictive in that a proportionate reduction in *all* inputs warrants an increase in *all* outputs. We now redefine (weak) congestion by relaxing such stringent requirements in the following definition.

Definition 12.2 (Weak congestion) A DMU is (weakly) congested if it is strongly efficient with respect to P_{convex} and there exist an activity in P_{convex} that uses less resources in one or more inputs for making more products in one or more outputs.

Notice that strong congestion implies weak congestion but not vice versa, and that, in a single input and a single output case, there is no distinction between strong and weak congestions. Weak but not strong congestion occurs only for the case with more than two inputs or two outputs.

In order to deal with this less restrictive definition, we present here an another scheme.

First, we assume that the concerned DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is strongly efficient in P_{convex} as defined in (12.25). Thus, the program [Congestion-strong] has ($\phi^* = 1, \mathbf{q}^{+*} = \mathbf{0}$) for every optimal solution. For this DMU $(\mathbf{x}_o, \mathbf{y}_o)$, we solve the following linear program with variables $\boldsymbol{\lambda}, \mathbf{t}^-$ and \mathbf{t}^+ .

$$\begin{aligned} [\text{Congestion-weak}] \quad & \max \frac{1}{s} \sum_{r=1}^s \frac{t_r^+}{y_{ro}} \\ \text{subject to} \quad & \mathbf{x}_o = X\boldsymbol{\lambda} + \mathbf{t}^- \\ & \mathbf{y}_o = Y\boldsymbol{\lambda} - \mathbf{t}^+ \\ & \mathbf{e}\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{t}^- \geq \mathbf{0}, \mathbf{t}^+ \geq \mathbf{0}. \end{aligned} \tag{12.63}$$

To solve [Congestion-weak], we employ a two-stage process: first, we maximize the objective function in (12.63), and then we maximize $\sum_{i=1}^m t_i^- / x_{io}$, while keeping the objective value of (12.63) at the optimum level. Symbolically, we utilize the objective function of form

$$\max \frac{1}{s} \sum_{r=1}^s \frac{t_r^+}{y_{ro}} + \varepsilon \frac{1}{m} \sum_{i=1}^m \frac{t_i^-}{x_{io}} \tag{12.64}$$

for the two-stage evaluation process, where ε is a non-Archimedean element smaller than any positive real number.

Let an optimal solution vector be $(\lambda^*, t^{-*}, t^{+*})$. We have now two cases:

Case 1. $t^{+*} = \mathbf{0}$.

In this case, no congestion is observed in activity (x_o, y_o) , since a decrease in inputs cannot increase any outputs.

Case 2. $t^{+*} \neq \mathbf{0}$.

In this case, t^{-*} is also not zero, since activity (x_o, y_o) is strongly efficient in P_{convex} . Thus, we identify (weak) congestion in activity (x_o, y_o) .

The objective function form in [Congestion-weak] is an output-oriented version of the Slacks-based Measure (SBM) that was introduced in Tone (2001).⁶ It is units invariant.

Henceforth, we deal with the Case 2, i.e., $t^{+*} \neq \mathbf{0}$, $t^{-*} \neq \mathbf{0}$. Based on the optimal solution vector $(\lambda^*, t^{-*}, t^{+*})$, we define \hat{x}_o and \hat{y}_o as

$$\hat{x}_o = X\lambda = x_o - t^{-*} \quad (12.65)$$

$$\hat{y}_o = Y\lambda = y_o + t^{+*}. \quad (12.66)$$

(\hat{x}_o, \hat{y}_o) is an improved (less congested) activity than (x_o, y_o) .

As a proxy measure for scale elasticity, we propose the following formula. First, we define an approximation to the marginal production rate (MPR) as

$$MPR = -\frac{1}{\bar{s}} \sum_{r=1}^s \frac{t_r^{+*}}{y_{ro}} \Big/ \frac{1}{\bar{m}} \sum_{i=1}^m \frac{t_i^{-*}}{x_{io}}, \quad (12.67)$$

where \bar{s} and \bar{m} are the numbers of positive t_r^{+*} ($r = 1, \dots, s$) and positive t_i^{-*} ($i = 1, \dots, m$), respectively. The average production rate (APR) is defined as

$$APR = \frac{1}{s} \sum_{r=1}^s \frac{y_{ro}}{y_o} \Big/ \frac{1}{m} \sum_{i=1}^m \frac{x_{io}}{x_o} = 1. \quad (12.68)$$

Thus, we have the following approximation measure for DSE, which is given by

$$DSE = \frac{MPR}{APR} = -\frac{1}{\bar{s}} \sum_{r=1}^s \frac{t_r^{+*}}{y_{ro}} \Big/ \frac{1}{\bar{m}} \sum_{i=1}^m \frac{t_i^{-*}}{x_{io}}. \quad (12.69)$$

This can be interpreted as the ratio of the average improvement in outputs to the average reduction in inputs. We also note here that the negative DSE value for any activity indicates that congestion is present in that activity.

We have the following theorem regarding the status of weak congestion and inefficiency in the BCC-O model.

Theorem 12.3 Suppose that the DMU (x_o, y_o) is efficient with respect to the P_{convex} . Then, it is weakly congested if and only if it has $\theta_{BCC}^* > 1$ or ($\theta_{BCC}^* = 1$ and $s^{+*} \neq \mathbf{0}$) by the model [BCC-O] in (12.3).

Proof: The status of weak congestion has two alternatives: (A) (weakly) Congested and (B) Not (weakly) congested, while the optimal solution for the [BCC-O] model has three alternatives: (a) $\theta_{BCC}^* > 1$, (b) $\theta_{BCC}^* = 1$ and $s^{+*} \neq \mathbf{0}$, and (c) $\theta_{BCC}^* = 1$ and $s^{+*} = \mathbf{0}$. We demonstrate that (A) corresponds to (a) or (b), and (B) corresponds to (c).

In case of (a), the second constraint in (12.3) can be transformed into:

$$\mathbf{y}_o = X\boldsymbol{\lambda}^* - (\theta_{BCC}^* - 1)\mathbf{y}_o - s^{+*}.$$

Since $\theta_{BCC}^* > 1$, it holds that $(\theta_{BCC}^* - 1)\mathbf{y}_o + s^{+*} > \mathbf{0}$ and hence the weak congestion model [Congestion-weak] has a positive objective function value, resulting in (A) (Case 2 above). In case of (b), s^{-*} must be non-zero, as otherwise the DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is not efficient with respect to the P_{convex} . Hence, it is weakly congested by definition and therefore (b) corresponds to (A). In case of (c), no congestion is identified, and (c) corresponds to (B). This concludes the proof. \square

Regarding the improved DMU $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$, we have the following theorem:

Theorem 12.4 *The improved DMU $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ defined by (12.65) and (12.66) is not weakly congested.*

Proof: The congestion status of the $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ is evaluated by the program below:

$$\begin{aligned} \max \quad & \frac{1}{s} \sum_{r=1}^s \frac{\tau_r^+}{y_{ro}} \\ \text{subject to} \quad & \hat{\mathbf{x}}_o = X\boldsymbol{\mu} + \boldsymbol{\tau}^- \\ & \hat{\mathbf{y}}_o = Y\boldsymbol{\mu} - \boldsymbol{\tau}^+ \\ & \mathbf{e}\boldsymbol{\mu} = 1 \\ & \boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\tau}^- \geq \mathbf{0}, \boldsymbol{\tau}^+ \geq \mathbf{0}. \end{aligned} \tag{12.70}$$

If it has an optimal solution $(\boldsymbol{\mu}^*, \boldsymbol{\tau}^{-*}, \boldsymbol{\tau}^{+*})$ with $\boldsymbol{\tau}^{+*} \neq \mathbf{0}$, then, referring to (12.65) and (12.66), it holds that

$$\begin{aligned} \mathbf{x}_o &= X\boldsymbol{\mu}^* + \boldsymbol{t}^{-*} + \boldsymbol{\tau}^{-*} \\ \mathbf{y}_o &= Y\boldsymbol{\mu}^* - \boldsymbol{t}^{+*} - \boldsymbol{\tau}^{+*}. \end{aligned}$$

Since $\boldsymbol{t}^{+*} + \boldsymbol{\tau}^{+*} \geq \boldsymbol{t}^{+*}$ and $\boldsymbol{t}^{+*} + \boldsymbol{\tau}^{+*} \neq \boldsymbol{t}^{+*}$, this contradicts the maximality of \boldsymbol{t}^{+*} for [Congestion-weak]. Thus, we have $\boldsymbol{\tau}^{+*} = \mathbf{0}$ and hence $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ is not weakly congested. \square

If an increase in some special outputs is required, the weights on the respective output slacks t_r^+ in the objective function (12.63) can be assigned in the following manner:

$$\max \frac{1}{s} \sum_{r=1}^s \frac{w_r t_r^+}{y_{ro}},$$

where the weight $w_r \geq 0$ (with $\sum_{r=1}^s w_r = s$) is the weight assigned to the output r .

12.3.3 Summary of Degree of Scale Economies and Congestion

We summarize here the procedure for finding out the degree of scale economies (diseconomies) and congestion (strong and weak) for the DMU $(\mathbf{x}_o, \mathbf{y}_o)$. We assume that $(\mathbf{x}_o, \mathbf{y}_o)$ is P_{convex} -efficient, i.e., $\phi^* = 1$ and $\mathbf{q}^{+*} = \mathbf{0}$ for every optimal solution for [Congestion-strong] (12.26). If not, we project $(\mathbf{x}_o, \mathbf{y}_o)$ onto the P_{convex} frontier by the formulas (12.27) and (12.28), and apply the following procedure to the projected DMU. In this case, the original DMU $(\mathbf{x}_o, \mathbf{y}_o)$ is technically inefficient and the projected DMU may be BCC-efficient, technically inefficient or congested.

Step 1. First, we solve [BCC-O] model (12.3).

- (A) If $\theta^* = 1$, $s^{-*} = \mathbf{0}$ and $s^{+*} = \mathbf{0}$, then $(\mathbf{x}_o, \mathbf{y}_o)$ is *BCC-efficient* and *not congested*. Further solve $\bar{\rho}$ and $\underline{\rho}$ in (12.24), and compute the scale elasticity (DSE) by taking their average. Stop.
- (B) If $\theta^* = 1$, $s^{-*} \neq \mathbf{0}$ and $s^{+*} = \mathbf{0}$, then $(\mathbf{x}_o, \mathbf{y}_o)$ is *technically inefficient*. Stop.
- (C) If $(\theta^* = 1 \text{ and } s^{+*} \neq \mathbf{0})$ or $\theta^* > 1$, then $(\mathbf{x}_o, \mathbf{y}_o)$ displays *congestion* by Theorem 3. Go to Step 2.

Step 2. Calculate the upper bound $\bar{\rho}$ of scale elasticity (12.39) by solving [Upper-w] (12.34)-(12.38).

(Strong Congestion) If $\bar{\rho} < 0$, then $(\mathbf{x}_o, \mathbf{y}_o)$ is *strongly congested*. Obtain $\underline{\rho}$ by solving the minimization problem (12.34)-(12.38). Combine $\bar{\rho}$ and $\underline{\rho}$ to compute DSE.

(Weak Congestion) If $\bar{\rho} \geq 0$, then $(\mathbf{x}_o, \mathbf{y}_o)$ is *weakly* but *not strongly* congested. Solve [Congestion-weak] (12.63) and calculate DSE by (12.69).

We note that, in the congested case, the improved activity is found respectively by (12.47) for the strongly congested DMU, and by (12.65) and (12.66) for the weakly congested DMU. However, since strong congestion also implies weak congestion, we can use (12.65) and (12.66) for the strongly congested DMU as well.

To compute DSE in the VRS environment, we can apply the procedure (12.19)-(12.23) for the BCC-O projected DMU, and then decide the nature of DSE via (12.24). In this case, no congestion is recognized.

12.4 ILLUSTRATIVE EXAMPLES

We demonstrate here the occurrence of congestion, the difference between congestion and technical inefficiency, and the distinction between weak and strong congestions with simple examples.

12.4.1 Degree of Scale Economies and Strong Congestion

We illustrate now the DSE and congestion with the help of input-output data of seven DMUs of Figure 2. As noted in Section 12.3.2, in this single input/output

case, there is no such distinction between strong and weak congestions. We report the DSE results of [BCC-O] and [Congestion-strong] models in Table 12.1.

Table 12.1. Example 1

DMU	Data			[BCC-O]		[Congestion-strong]		
	Input	Output	θ^*	$\bar{\rho}$	$\underline{\rho}$	RTS	$\bar{\rho}$	$\underline{\rho}$
A	2	1	1	∞	4	IRS		
B	3	3	1	2	1	CRS		
C	4	4	1	1	0.5	CRS		
D	6	5	1	0.6	0	DRS		
E	7	5	1					
F	8	4	1.25				-2	-4
G	9	2	2.5				-9	$-\infty$

From Figure 12.2, we can see that all DMUs are P_{convex} -efficient. So we applied the [BCC-O] model to the data set and found DMUs A, B, C and D being BCC-efficient and hence *not congested* by the rule (A) in Section 3.3. Their upper (lower) scale elasticity $\bar{\rho}$ ($\underline{\rho}$) is calculated by (12.24). From $\bar{\rho}$ and $\underline{\rho}$, we identified DMU A operating under IRS, DMUs B and C under CRS, and DMU D under DRS. DMU E has $\theta^* = 1$, $s^{-*} = 1$ and $s^{+*} = 0$, and hence, by the rule (B), E is *technically inefficient*. DMUs F and G have $\theta_F^* = 1.25$ and $\theta_G^* = 2.5$, and hence, by the rule (C), they are *congested*. Then we go to Step 2 to evaluate the scale elasticity of DMUs F and G. We found the upper elasticity $\bar{\rho}$ via (12.34)-(12.39), and the lower elasticity $\underline{\rho}$ using a similar scheme (by replacing max with min). DMUs F and G are found strongly congested with the degree of scale diseconomies score respectively at best -2 and -9.

12.4.2 Weak vs. Strong Congestion

We consider here four DMUs A, B, C and D, each uses two inputs (x_1 and x_2) to produce two outputs (y_1 and y_2). Their input and output data are displayed in Table 12.2.

Following the procedure described in Section 12.3.3, we first checked the P_{convex} efficiency score of these four DMUs, and found that they satisfied the efficiency condition: $\phi^* = 1$ and $q^{+*} = 0$ for every optimal solution for [Congestion-strong] (12.26). Then we solved the [BCC-O] model (12.3) and reported the results in Table 12.3.

DMUs A and B are BCC-efficient, since they have $\theta^* = 1$ and zero slacks. We then calculated their scale elasticity via (12.24) and found them exhibiting CRS. DMU C has $\theta^* = 1$ and $s_2^{+*} = 1$, and DMU D has $\theta^* = 2$, and hence they are all *congested* by the rule (C). Now we go to Step 2 to decide upon their status

Table 12.2. Example 2

DMU	Input		Output	
	x_1	x_2	y_1	y_2
A	1	1	1	1
B	2	2	2	2
C	2	3	2	1
D	3	3	1	1

Table 12.3. BCC-O Results

DMU	Score	Slacks				Scale elasticity		RTS
		s_1^{-*}	s_2^{-*}	s_1^{+*}	s_2^{+*}	$\bar{\rho}$	$\underline{\rho}$	
A	1	0	0	0	0	∞	1	CRS
B	1	0	0	0	0	1	0	CRS
C	1	0	1	0	1			
D	2	1	1	0	0			

concerning congestion. As can be seen from Table 12.4, ($\bar{\rho}_C = \infty$, $\underline{\rho}_C = -\infty$) and ($\bar{\rho}_D = -3$, $\underline{\rho}_D = -\infty$). DMU D is *strongly congested*, since $\bar{\rho}_D = -3 < 0$. Its DSE is -3 at best. DMU C is *weakly congested*, since $\bar{\rho}_C = \infty > 0$. We further evaluated its slacks t^{-*} and t^{+*} via (12.63) and (12.64), and reported them in Table 12.4. Using these values, we computed the DSE score of DMU C as $-(t_2^{+*}/y_{2C})/(t_2^{-*}/x_{2C}) = -(1/1)/(1/3) = -3$. We note that DMU C is found congested as compared with DMU B in both x_2 and y_2 , and hence DMU C can improve its performance by decreasing x_2 and by increasing y_2 , but not by all inputs/outputs, whereas DMU D has rooms for improving its performance not only by reducing all inputs but also by increasing all outputs. This clearly demonstrates the distinction between weak and strong congestions. Here both C and D are eventually projected to B, which is not congested at all.

12.5 SUMMARY OF CHAPTER 12

Investigation of scale elasticity for obtaining optimal scale of operations has significant bearings on recommending policy for restructuring any sector in a competitive economy. In recent years the BCC model has enjoyed widespread popularity in the non-parametric literature for computing the scale elasticity in production. However, there is a difficulty with the use of BCC model since

Table 12.4. Congestion Results

DMU	Scale elasticity		Slacks				Degree of scale Diseconomies	Status of congestion
	$\bar{\rho}$	$\underline{\rho}$	t_1^{-*}	t_2^{-*}	t_1^{+*}	t_2^{+*}		
C	∞	$-\infty$	0	1	0	1	-3	Weak
D	-3	$-\infty$	1	1	1	1	-3	Strong

this model does not directly take the congestion factors into consideration. When congestion is present in production, the BCC model overstates true scale elasticity estimates. It is of interest, therefore, to determine the impact of congestion on scale elasticity parameter of incorporating congested production factors into the model.

In this chapter, we have developed a new scheme to evaluate the scale elasticity in the presence of congestion within a unified framework.

12.6 NOTES AND SELECTED BIBLIOGRAPHY

Since the breakthrough by Banker, Charnes and Cooper (1984),⁷ the returns to scale issues under multiple input/output environments have been extensively studied in the framework of data envelopment analysis (DEA). We note that Banker *et al.*⁸ demonstrate that it is possible to compute exact scale elasticities by using the multiplicative model without using bounds or derivatives.

The treatment of congestion within the DEA framework has received considerable attention in the recent literature. After the concept of congestion was first introduced in the paper by Färe and Svensson (1980),⁹ and subsequently was given operationally implementable form by Färe *et al.* (1985)¹⁰ and Cooper *et al.* (1996, 2000),¹¹ there has been found a growing interest in congestion in a number of application areas, e.g., congestion in Chinese production by Brockett *et al.* (1998),¹² congestion in US teaching hospitals by Grosskopf *et al.* (2001).¹³ Just as the DEA literature (Färe *et al.*, 1986, 1988,¹⁴ Banker *et al.*, 1996a,b,¹⁵ Førsund, 1996¹⁶ and Sueyoshi, 1997,¹⁷ 1999¹⁸) that addresses the evaluation of scale elasticity is sparse, the research on using DEA to evaluate scale elasticity in the presence of congestion is almost nil.

We note that P.L. Brockett *et al.* (2004)¹⁹ give the following definition of congestion: "Definition 1: Evidence that congestion is present in the performance of any DMU when an increase in one or more inputs is associated with decreases that are maximally possible in one or more outputs, and these decreases are not accompanied by improvements in other inputs or outputs." Here, the term "maximally possible" is needed to distinguish between technical inefficiency and congestion since the former like the latter may have output increases occur simultaneously with input decreases simultaneously. See the discussion of Figure 1 in the above article.

12.7 RELATED DEA-SOLVER MODELS FOR CHAPTER 12

Scale Elasticity-I(O) (Input(Output)-oriented Scale Elasticity model)

The BCC model classifies the returns to scale of DMUs qualitatively, i.e., *increasing*, *constant* and *decreasing*. This Scale Elasticity model identifies the same characteristics quantitatively. The data format is the same as the CCR (or BCC) model.

Congestion (Congestion model)

This model detects congested DMUs and finds the sources of congestion quantitatively. The data format is the same as the CCR model.

12.8 PROBLEM SUPPLEMENT FOR CHAPTER 12

Problem 12.1

Prove Theorem 12.1

Proof: Let us change the units of X and Y respectively by diagonal matrices consisting of positive constants $D_x \in R^{m \times m}$ and $D_y \in R^{s \times s}$ as

$$X' = D_x X \text{ and } Y' = D_y Y.$$

By this units change, [BCC-O] turns out as

$$[\text{BCC-O}'] \quad \max \theta \quad (12.71)$$

$$\text{subject to} \quad \mathbf{x}'_o = X' \boldsymbol{\lambda} + \mathbf{s}^- \quad (12.72)$$

$$\theta \mathbf{y}'_o = Y' \boldsymbol{\lambda} - \mathbf{s}^+ \quad (12.73)$$

$$\mathbf{e} \boldsymbol{\lambda} = 1$$

$$\mathbf{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}.$$

By multiplying D_x^{-1} and D_y^{-1} from the left to (12.72) and (12.73) respectively, we obtain the essentially same expression as those in [BCC-O]. Thus, [BCC-O] suffers no effect.

On the other hand, the [Dual] results in

$$[\text{Dual}'] \quad \min \mathbf{v} \mathbf{x}'_o - w \quad (12.74)$$

$$\text{subject to} \quad -\mathbf{v} X' + \mathbf{u} Y' + \mathbf{e} w \leq \mathbf{0} \quad (12.75)$$

$$\mathbf{u} \mathbf{y}'_o = 1 \quad (12.76)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, w : \text{free in sign}, \quad (12.77)$$

which is equivalent to following model:

$$\min \mathbf{v} D_x \mathbf{x}_o - w \quad (12.78)$$

$$\text{subject to} \quad -\mathbf{v} D_x X + \mathbf{u} D_y Y + \mathbf{e} w \leq \mathbf{0} \quad (12.79)$$

$$\mathbf{u} D_y \mathbf{y}_o = 1 \quad (12.80)$$

$$\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, w : \text{free in sign}. \quad (12.81)$$

Let us replace vD_x by v' and uD_y by u' . Then, the dual program becomes

$$\min v'x_o - w \quad (12.82)$$

$$\text{subject to} \quad -v'X + u'Y + ew \leq 0 \quad (12.83)$$

$$u'y_o = 1 \quad (12.84)$$

$$v' \geq 0, u' \geq 0, w : \text{free in sign.} \quad (12.85)$$

Thus, we have now the same program as that in the [Dual] (12.7)-(12.9). Invoking the duality theorem of linear programming theory leads us to conclude the proof. \square

Notes

1. K. Tone and B.K. Sahoo (2004), "Degree of Scale Economies and Congestion: A Unified DEA Approach," *European Journal of Operational Research* 158, pp.755-772
2. R. Frisch (1965). *Theory of Production* Dordrecht: D. Reidel Publishing Company.
3. W.J. Baumol, J.C. Panzar and R.D. Willig, 1988. *Contestable Markets and The Theory of Industry Structure*. Harcourt Brace Jovanovich, New York.
4. R.D. Banker, A. Charnes and W.W. Cooper (1984), "Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30, pp.1078-1092. R.D. Banker, R.M. Thrall (1992), "Estimation of Returns to Scale Using Data Envelopment Analysis," *European Journal of Operational Research* 62, pp.74-84.
5. W.W. Cooper, H. Deng, B.S. Gu, S.L. Li and R.M. Thrall (2001), "Using DEA to Improve the Management of Congestion in Chinese Industries (1981-1997)," *Socio-Economic Planning Sciences* 35, pp.227-242.
6. K. Tone (2001), "A Slacks-based Measure of Efficiency in Data Envelopment Analysis," *European Journal of Operational Research* 130, pp.498-509.
7. R.D. Banker, A. Charnes and W.W. Cooper (1984), "Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* 30, pp.1078-1092.
8. See pp. 62 ff. in Chapter 2 of *Handbook on Data Envelopment Analysis*, Kluwer Academic Publishers, 2004.
9. R. Färe and L. Svensson (1980), "Congestion of Factors of Production," *Econometrica* 48, pp.1745-1753.
10. R. Färe, S. Grosskopf and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production* Boston: Kluwer Nijhoff.
11. W.W. Cooper, R.G. Thompson and R.M. Thrall (1996), "Introduction: Extensions and New Developments in DEA," *Annals of Operations Research* 66, pp.3-45. W.W. Cooper, L.M. Seiford and J. Zhu (2000), "A Unified Additive Model Approach for Evaluating Inefficiency and Congestion," *Socio-Economic Planning Sciences* 34, pp.1-26.
12. P.L. Brockett, W.W. Cooper, H.C. Shin and Y. Wang (1998), "Inefficiency and Congestion in Chinese Production before and after the 1978 Economic Reform," *Socio-Economic Planning Sciences* 32, pp.1-20.
13. S. Grosskopf, D. Margaritis and V. Valdmanis (2001), "Costs of Teaching on Hospital Productivity," *Socio-Economic Planning Sciences* 35, pp.189-204.
14. R. Färe, S. Grosskopf and C.A.K. Lovell (1986), "Scale Economies and Duality," *Journal of Economics* 46, pp.175-182. R. Färe, S. Grosskopf, C.A.K. Lovell (1988), "Scale Elasticity and Scale Efficiency," *Journal of Institutional and Theoretical Economics* 144, pp.721-729.
15. R.D. Banker, I. Bardhan and W.W. Cooper (1996a), "A Note on Returns to Scale in DEA," *European Journal of Operational Research* 88, pp.583-585. R.D. Banker, H. Chang, W.W. Cooper (1996b), "Equivalence and Implementation of Alternative Methods for Determining Returns to Scale in Data Envelopment Analysis," *European Journal of Operational Research* 89, pp.473-481.

16. F.R. Førsund (1996), "On the Calculation of Scale Elasticity in DEA Models," *Journal of Productivity Analysis* 7, pp.283-302.
17. T. Sueyoshi (1999), "DEA Duality on Returns to Scale (RTS) in Production and Cost Analyses: An Occurrence of Multiple Solutions and Differences Between Production-based and Cost-based RTS Estimates," *Management Science* 45, pp.1593-1609.
18. T. Sueyoshi (1997), "Measuring Efficiencies and Returns to Scale of Nippon Telegraph & Telephone in Production and Cost Analyses," *Management Science* 43, pp.779-796.
19. W.W. Cooper, H. Deng, L.L. Golden and T.W. Ruefli (2004), "Using DEA to Identify and Manage Congestion," *Journal of Productivity Analysis* 22, 207-226.

13 UNDESIRABLE OUTPUTS MODELS

13.1 INTRODUCTION

In accordance with the global environmental conservation awareness, undesirable outputs of productions and social activities, e.g., air pollutants and hazardous wastes, are being increasingly recognized as dangerous and undesirable. Thus, development of technologies with less undesirable outputs is an important subject of concern in every area of production. DEA usually assumes that producing more outputs relative to less input resources is a criterion of efficiency. In the presence of undesirable outputs, however, technologies with more good (desirable) outputs and less bad (undesirable) outputs relative to less input resources should be recognized as efficient.

This chapter deals with this environmental efficiency problem by applying a slacks-based measure of efficiency (SBM) introduced in Section 4.4 of Chapter 4. The SBM is non-radial and non-oriented, and utilizes input and output slacks directly in producing an efficiency measure. In this chapter, the SBM is modified so as to take undesirable outputs into account.

This chapter unfolds as follows. Section 2 extends the SBM to the case including undesirable outputs along with an economic interpretation of the model that enhances the rationale of the method. Extensions to various returns to scale assumptions and weight restriction are discussed, too. Then, in Section 3, we introduce a separable and non-separable inputs/outputs model that is

represented as an extension of the hybrid model (Hybrid) in Section 4.5 of Chapter 4. Illustrative examples are presented in Section 4.

13.2 AN SBM WITH UNDESIRABLE OUTPUTS

Suppose that there are n DMUs (decision making units) each having three factors: inputs, good outputs and bad (undesirable) outputs, as represented by three vectors $\mathbf{x} \in R^m$, $\mathbf{y}^g \in R^{s_1}$ and $\mathbf{y}^b \in R^{s_2}$, respectively. We define the matrices X , Y^g and Y^b as follows. $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in R^{m \times n}$, $Y^g = [\mathbf{y}_1^g, \dots, \mathbf{y}_n^g] \in R^{s_1 \times n}$, and $Y^b = [\mathbf{y}_1^b, \dots, \mathbf{y}_n^b] \in R^{s_2 \times n}$. We assume $X > 0$, $Y^g > 0$ and $Y^b > 0$.

The production possibility set (P) is defined by

$$P = \{(\mathbf{x}, \mathbf{y}^g, \mathbf{y}^b) | \mathbf{x} \geq X\boldsymbol{\lambda}, \mathbf{y}^g \leq Y^g\boldsymbol{\lambda}, \mathbf{y}^b \geq Y^b\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq 0\}, \quad (13.1)$$

where $\boldsymbol{\lambda} \in R^n$ is the intensity vector. Notice that the above definition corresponds to the constant returns to scale technology.

Definition 13.1 (Efficient DMU) A DMU_o $(\mathbf{x}_o, \mathbf{y}_o^g, \mathbf{y}_o^b)$ is efficient in the presence of undesirable outputs if there is no vector $(\mathbf{x}, \mathbf{y}^g, \mathbf{y}^b) \in P$ such that $\mathbf{x}_o \geq \mathbf{x}$, $\mathbf{y}_o^g \leq \mathbf{y}^g$ and $\mathbf{y}_o^b \geq \mathbf{y}^b$ with at least one strict inequality.

13.2.1 An Undesirable Output Model

In accordance with this definition, we modify the SBM as follows.

$$\text{[SBM-Undesirable]} \quad \rho^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s_1 + s_2} \left(\sum_{r=1}^{s_1} \frac{s_r^g}{y_{ro}^g} + \sum_{r=1}^{s_2} \frac{s_r^b}{y_{ro}^b} \right)} \quad (13.2)$$

$$\text{subject to } \mathbf{x}_o = X\boldsymbol{\lambda} + \mathbf{s}^- \quad (13.3)$$

$$\mathbf{y}_o^g = Y^g\boldsymbol{\lambda} - \mathbf{s}^g \quad (13.4)$$

$$\mathbf{y}_o^b = Y^b\boldsymbol{\lambda} + \mathbf{s}^b \quad (13.5)$$

$$\mathbf{s}^- \geq 0, \mathbf{s}^g \geq 0, \mathbf{s}^b \geq 0, \boldsymbol{\lambda} \geq 0.$$

The vectors $\mathbf{s}^- \in R^m$ and $\mathbf{s}^b \in R^{s_2}$ correspond to excesses in inputs and bad outputs, respectively, while $\mathbf{s}^g \in R^{s_1}$ expresses shortages in good outputs. The objective function (13.2) is strictly decreasing with respect to s_i^- ($\forall i$), s_r^g ($\forall r$) and s_r^b ($\forall r$) and the objective value satisfies $0 < \rho^* \leq 1$. Let an optimal solution of the above program be $(\boldsymbol{\lambda}^*, \mathbf{s}^{-*}, \mathbf{s}^{g*}, \mathbf{s}^{b*})$. Then, we have:

Theorem 13.1 The DMU_o is efficient in the presence of undesirable outputs if and only if $\rho^* = 1$, i.e., $\mathbf{s}^{-*} = \mathbf{0}$, $\mathbf{s}^{g*} = \mathbf{0}$ and $\mathbf{s}^{b*} = \mathbf{0}$.

If the DMU_o is inefficient, i.e., $\rho^* < 1$, it can be improved and become efficient by deleting the excesses in inputs and bad outputs, and augmenting the shortfalls in good outputs by the following SBM-projection:

$$\widehat{\mathbf{x}}_o \leftarrow \mathbf{x}_o - \mathbf{s}^{-*} \quad (13.6)$$

$$\hat{\mathbf{y}}_o^g \leftarrow \mathbf{y}_o^g + s^{g*} \quad (13.7)$$

$$\hat{\mathbf{y}}_o^b \leftarrow \mathbf{y}_o^b - s^{b*} \quad (13.8)$$

We notice that, in the objective function (13.2), the terms s_j^-/x_{io} and s_r^b/y_{ro}^b are bounded by 1 via the constraints (13.3) and (13.5). However, the term s_r^g/y_{ro}^g is unbounded as far as the constraint (13.4) is concerned. We can put an upper bound to this term if such a constraint is needed in real situations, e.g.,

$$\frac{s_r^g}{y_{ro}^g} \leq M_r, \quad (13.9)$$

where M_r designates the upper bound of the growth rate of y_{ro}^g .

Using the transformation by Charnes and Cooper (1962),¹ we have an equivalent linear program in t , Λ , \mathbf{S}^- , \mathbf{S}^g and \mathbf{S}^b as displayed below.

$$[\text{LP}] \quad \tau^* = \min t - \frac{1}{m} \sum_{i=1}^m \frac{S_i^-}{x_{io}} \quad (13.10)$$

$$\text{subject to} \quad 1 = t + \frac{1}{s_1 + s_2} \left(\sum_{r=1}^{s_1} \frac{S_r^g}{y_{ro}^g} + \sum_{r=1}^{s_2} \frac{S_r^b}{y_{ro}^b} \right) \quad (13.11)$$

$$\mathbf{x}_o t = X\Lambda + \mathbf{S}^- \quad (13.12)$$

$$\mathbf{y}_o^g t = Y^g \Lambda - \mathbf{S}^g \quad (13.13)$$

$$\mathbf{y}_o^b t = Y^b \Lambda + \mathbf{S}^b \quad (13.14)$$

$$\mathbf{S}^- \geq \mathbf{0}, \quad \mathbf{S}^g \geq \mathbf{0}, \quad \mathbf{S}^b \geq \mathbf{0}, \quad \Lambda \geq \mathbf{0}, \quad t > 0. \quad (13.15)$$

Let an optimal solution of [LP] be $(t^*, \Lambda^*, \mathbf{S}^{-*}, \mathbf{S}^{g*}, \mathbf{S}^{b*})$. Then we have an optimal solution of [SBM-Undesirable] as defined by

$$\rho^* = \tau^*, \quad \lambda^* = \Lambda^*/t^*, \quad s^{-*} = \mathbf{S}^{-*}/t^*, \quad s^{g*} = \mathbf{S}^{g*}/t^*, \quad s^{b*} = s^{b*}/t^*. \quad (13.16)$$

The existence of $(t^*, \Lambda^*, \mathbf{S}^{-*}, \mathbf{S}^{g*}, \mathbf{S}^{b*})$ with $t^* > 0$ is guaranteed by [LP].

13.2.2 Dual Interpretations

We have the dual program of [LP] as follows:

$$\xi^* = \max \xi \quad (13.17)$$

$$\text{subject to} \quad \xi + \mathbf{v}\mathbf{x}_o - \mathbf{u}^g \mathbf{y}_o^g + \mathbf{u}^b \mathbf{y}_o^b = 1 \quad (13.18)$$

$$-\mathbf{v}X + \mathbf{u}^g Y^g - \mathbf{u}^b Y^b \leq \mathbf{0}$$

$$\mathbf{v} \geq \frac{1}{m}[1/\mathbf{x}_o]$$

$$\mathbf{u}^g \geq \frac{\xi}{s}[1/\mathbf{y}_o^g]$$

$$\mathbf{u}^b \geq \frac{\xi}{s}[1/\mathbf{y}_o^b],$$

where $s = s_1 + s_2$ and the notation $[1/x_o]$ stands for the row vector $(1/x_{1o}, \dots, 1/x_{mo})$.

The dual variable vectors $v \in R^m$, $u^g \in R^{s_1}$ and $u^b \in R^{s_2}$ correspond to the constraints (13.12), (13.13) and (13.14), respectively. By eliminating ξ , this program is equivalent to the following:

$$[\text{DualLP}] \quad \max \quad u^g y_o^g - vx_o - u^b y_o^b \quad (13.19)$$

$$\text{subject to} \quad u^g Y^g - vX - u^b Y^b \leq 0 \quad (13.20)$$

$$v \geq \frac{1}{m} [1/x_o] \quad (13.21)$$

$$u^g \geq \frac{1 + u^g y_o^g - vx_o - u^b y_o^b}{s} [1/y_o^g] \quad (13.22)$$

$$u^b \geq \frac{1 + u^g y_o^g - vx_o - u^b y_o^b}{s} [1/y_o^b]. \quad (13.23)$$

The dual variables v and u^b can be interpreted as the virtual prices (costs) of inputs and bad outputs, respectively, while u^g denotes the dual price of good outputs. The dual program aims at obtaining the optimal virtual costs and prices for DMU_o so that the profit $u^g y^g - vx - u^b y^b$ does not exceed zero for every DMU (including DMU_o) and maximizes the profit $u^g y_o^g - vx_o - u^b y_o^b$ for the DMU_o concerned. Apparently, the optimal profit is at best zero and hence $\xi^* = 1$ for the SBM efficient DMUs.

13.2.3 Returns-to-scale (RTS) Issues

Although we discussed our bad outputs model under the constant returns to scale (CRS) assumption, we can incorporate other RTS by adding the following constraint to [SBM-Undesirable] and hence to the definition of the production possibility set P ,

$$L \leq e\lambda \leq U, \quad (13.24)$$

where $e = (1, \dots, 1) \in R^n$ and, $L(\leq 1)$ and $U(\geq 1)$ are respectively the lower and upper bounds to the intensity λ .

The cases ($L = 1$, $U = 1$), ($L = 0$, $U = 1$) and ($L = 1$, $U = \infty$) correspond to the variable (VRS), the decreasing (DRS) and the increasing (IRS) RTS, respectively.

The definition of the efficiency status is the same as described in [Definition 1] and Theorem 1 holds in these cases, too.

13.2.4 Imposing Weights to Inputs and/or Outputs

If putting preference (or importance) on input/output items is requested, we can impose weights to the objective function in (13.2) as follows:

$$[\text{BadOutput-Weight}] \quad \rho^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{w_i^- s_i^-}{x_{io}}}{1 + \frac{1}{s_1+s_2} \left(\sum_{r=1}^{s_1} \frac{w_r^g s_r^g}{y_{ro}^g} + \sum_{r=1}^{s_2} \frac{w_r^b s_r^b}{y_{ro}^b} \right)}, \quad (13.25)$$

where w_i , w_r^g and w_r^b are the weights to the input i , the desirable output r , and the undesirable output r , respectively, and $\sum_{i=1}^m w_i^- = m$, $w_i^- \geq 0 (\forall i)$, $\sum_{r=1}^{s_1} w_r^g + \sum_{r=1}^{s_2} w_r^b = s_1 + s_2$, $w_r^g \geq 0 (\forall r)$, $w_r^b \geq 0 (\forall r)$.

13.3 NON-SEPARABLE ‘GOOD’ AND ‘BAD’ OUTPUT MODEL

It is often observed that certain ‘bad’ outputs are not separable from the corresponding ‘good’ outputs. In each case, reducing bad outputs is inevitably accompanied by reduction in good outputs. Furthermore, it often occurs that a certain bad output is closely related (non-separable) with a certain input. For example, in electric utility industries, emissions of Nitrogen Oxides (NO_x) and Sulfur Dioxides (SO_2) (bad outputs) are proportional to the fuel consumption in inputs. In this section, we discuss this non-separable case. For this, we decompose the set of good and bad outputs (Y^g, Y^b) into (Y^{Sg}) and (Y^{NSg}, Y^{NSb}) , where $Y^{Sg} \in R^{s_{11} \times n}$ and $(Y^{NSg} \in R^{s_{21} \times n}, Y^{NSb} \in R^{s_{22} \times n})$ denote the separable good outputs, and non-separable good and bad outputs, respectively. The set of input X is decomposed into (X^S, X^{NS}) where $X^S \in R^{m_1 \times n}$ and $X^{NS} \in R^{m_2 \times n}$ denote respectively the separable and non-separable inputs. For the separable outputs Y^{Sg} , we have the same structure of production as Y^g in P . However, the non-separable outputs (Y^{NSg}, Y^{NSb}) need another handling. A reduction of the bad outputs y^{NSb} is designated by αy^{NSb} with $0 \leq \alpha \leq 1$, which is accompanied by a proportionate reduction in the good outputs y^{NSg} as denoted by αy^{NSg} as well as in the non-separable input denoted by αx^{NS} .

Now, the new production possibility set P_{NS} under CRS is defined by:

$$\begin{aligned} P_{NS} = & \left\{ (x^S, x^{NS}, y^{Sg}, y^{NSg}, y^{NSb}) \mid \right. \\ & x^S \geq X^S \lambda, x^{NS} \geq X^{NS} \lambda, y^{Sg} \leq Y^{Sg} \lambda, \\ & y^{NSg} \leq Y^{NSg} \lambda, y^{NSb} \geq Y^{NSb} \lambda, \lambda \geq 0 \left. \right\}. \end{aligned} \quad (13.26)$$

Basically this definition is a natural extension of P in (13.1). We alter the definition of the efficiency status in the non-separable case as follows:

Definition 13.2 (NS-efficient) A DMU_o $(x_o^S, x_o^{NS}, y_o^{Sg}, y_o^{NSg}, y_o^{NSb})$ is called NS-efficient if and only if

(1) for any α ($0 \leq \alpha < 1$), we have

$$(x_o^S, x_o^{NS}, y_o^{Sg}, \alpha y_o^{NSg}, \alpha y_o^{NSb}) \notin P_{NS},$$

and (2) there is no $(x^S, x^{NS}, y^{Sg}, y^{NSg}, y^{NSb}) \in P_{NS}$ such that

$$x_o^S \geq x^S, x_o^{NS} = x^{NS}, y_o^{Sg} \leq y^{Sg}, y_o^{NSg} = y^{NSg}, y_o^{NSb} = y^{NSb}$$

with at least one strict inequality.

An SBM with the non-separable inputs and outputs can be implemented by the program in $(\lambda, s^{S-}, s^{Sg}, \alpha)$ as below:

[SBM-NS]

$$\rho^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^{m_1} \frac{s_i^{S-}}{x_{io}} - \frac{m_2}{m}(1-\alpha)}{1 + \frac{1}{s} \left(\sum_{r=1}^{s_{11}} \frac{s_r^{Sg}}{y_{ro}^{Sg}} + (s_{21} + s_{22})(1-\alpha) \right)} \quad (13.27)$$

subject to

$$\mathbf{x}_o^S = X^S \boldsymbol{\lambda} + \mathbf{s}^{S-} \quad (13.28)$$

$$\alpha \mathbf{x}_o^{NS} \geq X^{NS} \boldsymbol{\lambda} \quad (13.29)$$

$$\mathbf{y}_o^{Sg} = Y^{Sg} \boldsymbol{\lambda} - \mathbf{s}^{Sg} \quad (13.30)$$

$$\alpha \mathbf{y}_o^{NSg} \leq Y^{NSg} \boldsymbol{\lambda} \quad (13.31)$$

$$\alpha \mathbf{y}_o^{NSb} \geq Y^{NSb} \boldsymbol{\lambda} \quad (13.32)$$

$$\mathbf{s}^{S-} \geq \mathbf{0}, \mathbf{s}^{Sg} \geq \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}, 0 \leq \alpha \leq 1,$$

where $m = m_1 + m_2$ and $s = s_{11} + s_{21} + s_{22}$.

This is a variant of the Hybrid model introduced in Section 4.5 of Chapter 4. The objective function is strictly monotone decreasing with respect to s_i^{S-} ($\forall i$), s_r^{Sg} ($\forall r$) and α . Let an optimal solution for [SBM-NS] be $(\rho^*, \boldsymbol{\lambda}^*, \mathbf{s}^{S-*}, \mathbf{s}^{Sg*}, \alpha^*)$, then we have $0 < \rho^* \leq 1$ and the following theorem holds:

Theorem 13.2 *The DMU_o is non-separable (NS)-efficient if and only if $\rho^* = 1$, i.e., $\mathbf{s}^{S-*} = \mathbf{0}$, $\mathbf{s}^{Sg*} = \mathbf{0}$ and $\alpha^* = 1$.*

If the DMU_o is NS-inefficient, i.e., $\rho^* < 1$, it can be improved and become NS-efficient by the following NS-projection:

$$\hat{\mathbf{x}}_o^S \leftarrow \mathbf{x}_o^S - \mathbf{s}^{S-*} \quad (13.33)$$

$$\hat{\mathbf{x}}_o^{NS} \leftarrow \alpha^* \mathbf{x}_o^{NS} \quad (13.34)$$

$$\hat{\mathbf{y}}_o^{Sg} \leftarrow \mathbf{y}_o^{Sg} + \mathbf{s}^{Sg*} \quad (13.35)$$

$$\hat{\mathbf{y}}_o^{NSg} \leftarrow \alpha^* \mathbf{y}_o^{NSg} \quad (13.36)$$

$$\hat{\mathbf{y}}_o^{NSb} \leftarrow \alpha^* \mathbf{y}_o^{NSb}. \quad (13.37)$$

It should be noted that, from (13.31) and (13.32), it holds that

$$\mathbf{s}^{NS-*} \equiv -\alpha^* \mathbf{x}_o^{NS} + X^{NS} \boldsymbol{\lambda} \geq \mathbf{0} \quad (13.38)$$

$$\mathbf{s}^{NSg*} \equiv -\alpha^* \mathbf{y}_o^{NSg} + Y^{NSg} \boldsymbol{\lambda}^* \geq \mathbf{0} \quad (13.39)$$

$$\mathbf{s}^{NSb*} \equiv \alpha^* \mathbf{y}_o^{NSb} - Y^{NSb} \boldsymbol{\lambda}^* \geq \mathbf{0}. \quad (13.40)$$

This means that some of slacks in non-separable inputs and outputs may remain positive even after the projection and that these slacks, if any, are not accounted for in the NS-efficiency score, since we assume the proportionate reduction (α^*) in these outputs. Thus, we apply the SBM for the separable outputs, whereas we employ the radial approach for the non-separable outputs.

In actual situations, it is often required that, in addition to the constraints (13.28)-(13.32), the total amount of good outputs should remain unchanged

and the expansion rate of separable good outputs should be bounded by an exogenous value. The former option is described as

$$\sum_{r=1}^{s_{11}} (y_{ro}^{Sg} + s_r^{Sg}) + \alpha \sum_{r=1}^{s_{21}} y_{ro}^{NSg} = \sum_{r=1}^{s_{11}} y_{ro}^{Sg} + \sum_{r=1}^{s_{21}} y_{ro}^{NSg}, \quad (13.41)$$

where we assume that the measurement units are the same among all good outputs. The latter condition can be expressed as

$$\frac{s_r^{Sg}}{s_g^{Sg}} \leq U, \quad (\forall r) \quad (13.42)$$

where U is the upper bound to the expansion rate for the separable good outputs.

Furthermore, it is reasonable that the slacks in non-separable (radial) bad outputs and non-separable inputs should affect the overall efficiency, since even the radial slacks are sources of inefficiency.

Summing all these requirements, we have the following model for evaluating overall efficiency.

[NS-Overall]

$$\rho^* = \min \frac{1 - \frac{1}{m} \sum_{i=1}^{m_1} \frac{s_i^{S-}}{x_{io}^S} - \frac{1}{m} \sum_{i=1}^{m_2} \frac{s_i^{NS-}}{x_{io}^{NS}} - \frac{m_2}{m} (1 - \alpha)}{1 + \frac{1}{s} \left(\sum_{r=1}^{s_{11}} \frac{s_r^{Sg}}{y_{ro}^{Sg}} + \sum_{r=1}^{s_{21}} \frac{s_r^{NSb}}{y_{ro}^{NSb}} + (s_{21} + s_{22})(1 - \alpha) \right)} \quad (13.43)$$

subject to

$$x_o^S = X^S \lambda + s^{S-} \quad (13.44)$$

$$\alpha x_o^{NS} = X^{NS} \lambda + s^{NS-} \quad (13.45)$$

$$y_o^{Sg} = Y^{Sg} \lambda - s^{Sg} \quad (13.46)$$

$$\alpha y_o^{NSg} \leq Y^{NSg} \lambda \quad (13.47)$$

$$\alpha y_o^{NSb} = Y^{NSb} \lambda + s^{NSb} \quad (13.48)$$

$$\sum_{r=1}^{s_{11}} (y_{ro}^{Sg} + s_r^{Sg}) + \alpha \sum_{r=1}^{s_{21}} y_{ro}^{NSg} = \sum_{r=1}^{s_{11}} y_{ro}^{Sg} + \sum_{r=1}^{s_{21}} y_{ro}^{NSg} \quad (13.49)$$

$$\frac{s_r^{Sg}}{s_g^{Sg}} \leq U \quad (\forall r) \quad (13.50)$$

$$s^{S-} \geq 0, \quad s^{NS-} \geq 0, \quad s^{Sg} \geq 0, \quad s^{NSb} \geq 0, \quad \lambda \geq 0, \quad 0 \leq \alpha \leq 1.$$

Similarly to the decomposition of inefficiency in Section 4.5.2 of Chapter 4, we can decompose this overall efficiency into respective inefficiencies as follows:

$$\rho^* = \frac{1 - \sum_{i=1}^{m_1} \alpha_{1i} - \sum_{i=1}^{m_2} \alpha_{2i}}{1 + \sum_{r=1}^{s_{11}} \beta_{1r} + \sum_{r=1}^{s_{21}} \beta_{2r} + \sum_{r=1}^{s_{22}} \beta_{3r}} \quad (13.51)$$

where

$$\alpha_{1i} = \frac{1}{m} \frac{s_i^{S-*}}{x_{io}^S} \quad (i = 1, \dots, m_1) \quad (\text{Separable Input}) \quad (13.52)$$

$$\alpha_{2i} = \frac{1}{m} (1 - \alpha^*) + \frac{1}{m} \frac{s_i^{NS-*}}{x_{io}^{NS}} \quad (i = 1, \dots, m_2) \quad (\text{Non-Separable Input}) \quad (13.53)$$

$$\beta_{1r} = \frac{1}{s} \frac{s_r^{Sg*}}{y_{ro}^{Sg}} \quad (r = 1, \dots, s_{11}) \quad (\text{Separable Good Output}) \quad (13.54)$$

$$\beta_{2r} = \frac{1}{s} (1 - \alpha^*) \quad (r = 1, \dots, s_{21}) \quad (\text{Non-Separable Good Output}) \quad (13.55)$$

$$\beta_{3r} = \frac{1}{s} (1 - \alpha^*) + \frac{1}{s} \frac{s_r^{NSb*}}{y_{ro}^{NSb}} \quad (r = 1, \dots, s_{22}) \quad (\text{Non-Separable Bad Output}) \quad (13.56)$$

13.4 ILLUSTRATIVE EXAMPLES

We present numerical examples that illustrate the aforementioned schemes.

13.4.1 Separable Bad Outputs Models

Table 13.1 exhibits a simple data set composed of a single input (with value=1) and two outputs: one desirable and one undesirable where ‘bad’ is assumed to be separable from ‘good’. We solved this problem by employing the weight

Table 13.1. Separable Bad Outputs Case: Data set

DMU	x	Output	
		Desirable y^{Good}	Undesirable y^{Bad}
A	1	1	1
B	1	2	1
C	1	6	2
D	1	8	4
E	1	9	7
F	1	5	2
G	1	4	3
H	1	6	4
I	1	4	6

selection scheme [BadOutput-Weight] (13.25) described in Section 2.4. We put three ratios of weights on ‘good’ and ‘bad’ outputs: (1:0.3), (1:1) and (1:3). Table 13.2 reports the results: the efficiency (ρ^*), and the projected good and bad outputs. DMUs B, C, D and E are efficient in this separable outputs

model. It is observed that, as the weight moves from good to bad, the emphasis of projection changes from enlargement of the good output to reduction of the bad output.

Table 13.2. Separable Bad Outputs Case: Results

ρ^*	Weights to Good and Bad Outputs								
	1:0.3		1:1		1:3				
	Projected	y^{Good}	Projected	y^{Good}	Projected	y^{Good}			
ρ^*	y^{Bad}		ρ^*	y^{Bad}	ρ^*	y^{Bad}			
A	0.565	2	1	0.667	2	1	0.8	2	1
B	1	-	-	1	-	-	1	-	-
C	1	-	-	1	-	-	1	-	-
D	1	-	-	1	-	-	1	-	-
E	1	-	-	1	-	-	1	-	-
F	0.867	6	2	0.909	6	2	0.914	5	1.75
G	0.634	7	3	0.706	6	2	0.727	6	2
H	0.796	8	4	0.80	6	2	0.727	6	2
I	0.527	8.67	6	0.60	8	4	0.615	6	2

13.4.2 An Example with Both Separable and Non-separable Inputs/Outputs

Table 13.3 displays inputs and outputs of five electric utilities, A to H, at years 1996 and 2000. For example, 1996A indicates Utility A at year 1996. We employ two inputs and three outputs as follows:

- Input:
 1. Total generation capacity (Separable)
 2. Fuel consumption (Non-separable)
- Output:
 1. Non-fossil power generation (Separable good)
 2. Fossil power generation (Non-separable good)
 3. Nitrogen oxides (NO_x emission) (Non-separable bad)

NO_x emission is an undesirable output of electric power industries which is non-separable from fossil generation (output) and fuel consumption (input). The purpose here is to compare the performance of 1996 DMUs with that of 2000 in terms of both good and bad outputs relative to the separable and non-separable inputs. We applied the [NS-Overall] model assuming the upper bound $U = 0.2$ of the expansion rate of the separable good output (non-fossil

Table 13.3. Non-Separable Inputs/Outputs Case: Data Set

DMU	Input Sep Capacity	Input Non-Sep Fuel	Output Sep(Good) Non-Fossil Gene	Output Non-Sep(Good) Fossil Gene	Output Non-Sep(Bad) NO _x
1996A	129	559	178	544	156
1996B	86	274	102	246	54
1996C	92	225	289	194	57
1996D	105	289	243	257	117
1996E	83	183	53	178	52
1996F	115	458	48	435	105
1996G	153	359	265	332	110
1996H	200	431	480	408	163
2000A	141	637	149	599	123
2000B	86	330	109	307	48
2000C	92	283	304	247	65
2000D	121	334	274	315	69
2000E	78	211	50	193	42
2000F	122	441	82	433	95
2000G	164	431	277	412	94
2000H	200	420	569	440	100

generation) and obtained the results displayed in Table 13.4. The efficient frontiers are formed by 1996A, 1996C, 2000A, 2000B, 2000C and 2000H. The other DMUs have slacks against these frontiers and their inefficiencies can be decomposed into each factor using the formulae (13.52)-(13.56).

From the results, we can see that the average overall efficiency (ρ^*) improves from 0.770 (1996) to 0.875 (2000) and all inefficiencies decrease from 1996 to 2000. Especially, NO_x inefficiency is the top in decrement. This indicates that these companies improved their efficiency in management as well as in environmental awareness. Tone and Tsutsui (2006)² discuss this subject in detail.

13.5 COMPARISONS WITH OTHER METHODS

Several authors have proposed efficiency measures in the presence of undesirable outputs. A conventional and traditional way to handle this problem is to shift undesirable outputs to inputs and to apply traditional DEA models to the data set. In the VRS environment, Seiford and Zhu (2002)³ proposed a method that first multiplies each undesirable output by -1 and then finds a proper translation vector to let all negative undesirable outputs be positive. Interestingly, Scheel (2001)⁴ pointed out that these two transformations (position

Table 13.4. Non-Separable Inputs/Outputs Case: Decomposition of Inefficiency

DMU	ρ^*	α^*	Decomposition of Inefficiency					
			Capacity α_1	Fuel α_2	Other Gene β_1	Fossil β_2	NO _x β_3	
1996A	1	1	0	0	0	0	0	0
1996B	0.7	0.865	0.087	0.067	0.067	0.045	0.096	
1996C	1	1	0	0	0	0	0	
1996D	0.621	0.720	0.028	0.140	0.053	0.093	0.192	
1996E	0.639	0.940	0.210	0.030	0.067	0.020	0.103	
1996F	0.891	0.978	0.018	0.011	0.067	0.007	0.016	
1996G	0.590	0.754	0.115	0.123	0.067	0.082	0.144	
1996H	0.715	0.858	0.060	0.071	0.014	0.047	0.153	
1996Average	0.770	0.889	0.065	0.055	0.042	0.037	0.088	
2000A	1	1	0	0	0	0	0	
2000B	1	1	0	0	0	0	0	
2000C	1	1	0	0	0	0	0	
2000D	0.760	0.856	0.016	0.109	0.055	0.048	0.048	
2000E	0.662	0.930	0.185	0.035	0.067	0.023	0.088	
2000F	0.887	0.962	0.012	0.019	0.067	0.013	0.013	
2000G	0.693	0.828	0.079	0.086	0.067	0.057	0.08	
2000H	1	1	0	0	0	0	0	
2000Average	0.875	0.947	0.037	0.031	0.032	0.018	0.029	
Ratio 1996 vs. 2000			57%	56%	76%	49%	33%	

change and translation) give the same efficient frontiers, although the Seiford and Zhu method is only valid under the VRS condition. However, resulting efficiency scores for inefficient DMUs are different by the model employed. Another conventional way is to invert the undesirable output value and treat it as a desirable one. This operation may cause the deformation of the efficient frontiers due to the non-linear transformation and hence gives a different identification of the efficiency status and efficiency score.

Färe *et al.* (1989)⁵ was the first paper to treat this subject systematically. They treat desirable and undesirable outputs asymmetrically, resulting in an enhanced hyperbolic output efficiency measure. This approach needs to solve a non-linear programming problem. As to the non-separable models, Scheel (2001) proposed a radial and output-oriented method, whereas Färe *et al.* (2003)⁶ developed a directional vector approach in output-orientation. The non-separable outputs models have different efficient frontiers than the separable outputs case.

We would like to consider this problem from another standpoint. Most (but not all) DEA models can be categorized into four classes: (1) radial and oriented, (2) radial and non-oriented, (3) non-radial and oriented, and (4) non-radial and non-oriented. Here, ‘radial’ means that a proportional reduction or enlargement of inputs/outputs is the main concern in measuring efficiency, while ‘oriented’ indicates input-oriented or output-oriented. Consequently, radial models neglect slacks and hence, when dealing with undesirable outputs, slacks in undesirable outputs are not accounted for in the efficiency measure. This is a crucial shortcoming of radial models. On the other hand, the major concern of input (output)-oriented models focuses on the input (output)-side efficiency, and output (input)-side is a minor subject in measuring efficiency. Thus, only the non-radial and non-oriented models can capture the whole aspects of efficiency. From this viewpoint, it should be noted that our [SBM-NS] and [NS-Overall] models can deal with inputs and outputs efficiency in the presence of both non-separable and separable outputs in a unified manner in any RTS environments.

13.6 SUMMARY OF CHAPTER 13

In accordance with the recent environmental movements, it may be required to evaluate the relative efficiency of enterprises within the framework that includes both desirable and undesirable outputs. In this chapter we have introduced a new SBM scheme for this purpose. The new scheme provides the full efficiency score 1 to a DMU if and only if the DMU is efficient and assigns a score less than 1 to inefficient DMUs depending on the relative magnitudes of slacks in inputs/outputs. It is strictly monotone decreasing with respect to an increase in any slacks. Furthermore, the proposed method processes the given data set as it is, i.e., neither translation nor inversion is required, and no position change (from output to input) is required.

This scheme can be extended to cope with co-existence of non-separable desirable and undesirable outputs as well as separable ones. It belongs to a non-radial and non-oriented index and has a reasonable LP dual interpretation as a virtual profit maximization engine.

A major concern of our society is how to better take into considerations both economic and ecological issues. For this, as Korhonen and Luptacik (2004)⁷ note, *we need new indicators to measure the economic performance of the firm and the national economy, which take into account environmental aspects as well*. We hope the new methodology developed in this chapter can contribute to this purpose.

13.7 RELATED DEA-SOLVER MODELS FOR CHAPTER 13

BadOutput-C (V, GRS) (The BadOutput model under constant (variable, general) returns-to-scale assumption). This code solves the weighted bad output model [BadOutput-Weight] in (13.25). See Appendix B.5.13 for data format.

NonSeparable-C (V, GRS) (The Non-separable input/output model under constant (variable, general) returns-to-scale assumption). This code solves the non-separable input/output model [NS-Overall] in (13.43). See Appendix B.5.14 for data format.

Notes

1. A. Charnes and W.W. Cooper (1962), "Programming with Linear Fractional Functionals," *Naval Research Logistics Quarterly* 15, pp.333-334.
2. K. Tone and M. Tsutsui (2006), "An Efficiency Measure of Goods and Bads in DEA and its Application to US Electric Utilities," presented at Asia Pacific Productivity Conference 2006, Korea.
3. L.M. Seiford and J. Zhu (2002), "Modeling Undesirable Factors in Efficiency Evaluation," *European Journal of Operational Research* 142, 16-20.
4. H. Scheel (2001), "Undesirable Outputs in Efficiency Valuations," *European Journal of Operational Research* 132, pp.400-410.
5. R. Färe, S. Grosskopf, C.A.K. Lovell and C. Pasurka (1989), "Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach," *The Review of Economics and Statistics* 71, pp.90-98.
6. R. Färe and S. Grosskopf (2003), "Modeling Undesirable Factors in Efficiency Evaluation: Comment," *European Journal of Operational Research* (forthcoming)
7. P.J. Korhonen and M. Luptacik (2004), "Eco-efficiency Analysis of Power Plants: An Extension of Data Envelopment Analysis," *European Journal of Operational Research* 154, 437-446.

14 ECONOMIES OF SCOPE AND CAPACITY UTILIZATION

14.1 INTRODUCTION

“Economies of scope” was introduced into the economics literature by Baumol, Panzar and Willig (1982)¹ for use in determining whether a firm might advantageously produce a variety of products and services or whether it would be better to spin some of them off for production by separate entities. Färe (1986, 1988),² extended Baumol *et al.* for use with DEA, and Grosskopf and Yaisawarng (1990)³ adapted this use of DEA for application to public service agencies in order to determine when some of the separate agencies delivering these services might be advantageously consolidated. See also Chapter 10 in Färe, Grosskopf and Lovell (1994).⁴

In the first part of this chapter we also use DEA to determine whether, and where, diversification is superior to specialization among the entities producing specified goods and services. Our model and methods differ from those of Färe, and Grosskopf and Yaisawarng, and thus widen and extend these uses of DEA. In the immediately following section we introduce the concept of economies of scope. We then show how this concept may be used to evaluate merger and acquisition possibilities after which we apply our models and concepts to an example involving electric power production and distribution. Finally, in the last part of this chapter, we turn to the topics of “capacity utilization” which Johansen (1968)⁵ defined as “the maximum amount that can be produced per unit time.”

14.2 ECONOMIES OF SCOPE

In this section we first introduce the concept of economies of scope using the case of a single input (cost) and two outputs where the two outputs belong to different categories, e.g., the amount of electric power generation and the number of customers served in electric utility companies. Diversified companies produce both, while specialized companies produce only one of them. Then we discuss the difference in efficiency between two business models which use the same category of inputs, e.g., capital, labor and materials to produce the same category of outputs, e.g., operating profits.

14.2.1 Definition

Baumol *et al.* (1982) define economies of scope to be present between two products (y_1, y_2) if the cost of producing both products by one firm is less than the cost of producing them separately in specialized firms. Mathematically, economies of scope is present if

$$C(y_1, y_2) < C_1(y_1, 0) + C_2(0, y_2), \quad (14.1)$$

where $C(y_1, y_2)$ is the cost of joint production by the diversified firm, and $C_1(y_1, 0)$ and $C_2(0, y_2)$ are the respective costs of production of y_1 and y_2 by two specialized firms. So, the degree of economies of scope (DES) for firm j is defined as

$$\text{DES}_j = \frac{C_1(y_1, 0) + C_2(0, y_2) - C(y_1, y_2)}{C(y_1, y_2)}. \quad (14.2)$$

$\text{DES}_j > 0$ implies that firm j exhibits economies of scope, $\text{DES}_j < 0$ implies diseconomies of scope, and $\text{DES}_j = 0$ means that the costs are additive in nature. Note that the non-zero outputs are the same on both sides of (14.1) so the evaluation is applicable only locally (i.e., in the vicinity of y_1 and y_2).

14.2.2 Checking for Economies of Scope

In order to handle this problem in the DEA framework, we utilize the following settings. There are n diversified DMUs, called group D, that produce both Products 1 and 2. For the diversified DMU j , we denote the outputs and input (cost) by $\mathbf{y}_j = (y_{1j}, y_{2j})$ and c_j ($j = 1, \dots, n$), respectively. We define the production possibility set of diversified firms under the variable returns-to-scale assumption as

$$P_{1,2} = \left\{ (c, \mathbf{y}) \mid c = \sum_{j=1}^n c_j \lambda_j, \mathbf{y} \leq \sum_{j=1}^n \mathbf{y}_j \lambda_j, \mathbf{e}\lambda = 1, \lambda \geq 0 \right\}, \quad (14.3)$$

so, $c_j \lambda_j$ = total cost of firm j to produce both products, where $\mathbf{y}_j \lambda_j = \begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} \lambda_j$ and c is the total cost of the firm to produce \mathbf{y} .

There are also p specialized firms, group S_1 , that produce only Product 1. This group produces output z_{1j} using input (cost) v_{1j} for $j = 1, \dots, p$.

The other specialized group, denoted by S_2 , produces Product 2 (z_{2j}) using cost (v_{2j}) for $j = 1, \dots, q$ where q is the number of firms in group S_2 . Their production possibility sets are denoted by P_1 and P_2 , respectively,

$$P_1 = \{(v, z_1) | v = \sum_{j=1}^p v_{1j} \lambda_j, z_1 \leq \sum_{j=1}^p z_{1j} \lambda_j, e\lambda = 1, \lambda \geq 0\}. \quad (14.4)$$

$$P_2 = \{(v, z_2) | v = \sum_{j=1}^q v_{2j} \lambda_j, z_2 \leq \sum_{j=1}^q z_{2j} \lambda_j, e\lambda = 1, \lambda \geq 0\}. \quad (14.5)$$

Figure 14.1 depicts the above input/output flow.

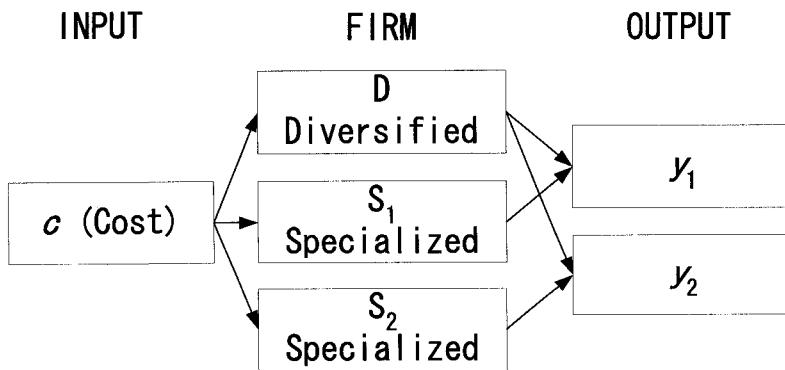


Figure 14.1. Diversified and Specialized Firms

For $\text{DMU } (c_o, y_{1o}, y_{2o}) \in D$, we solve the following LP for each DMU in order to eliminate any cost-inefficiency,

$$\begin{aligned} [\text{Cost}] \quad C(y_{1o}, y_{2o}) &= c_o^* = \min_{\theta, \lambda} c_o \theta & (14.6) \\ c_o \theta &= c_1 \lambda_1 + c_2 \lambda_2 + \dots + c_n \lambda_n \\ y_{io} &\leq y_{i1} \lambda_1 + y_{i2} \lambda_2 + \dots + y_{in} \lambda_n \quad (i = 1, 2) \\ 1 &= \lambda_1 + \lambda_2 + \dots + \lambda_n \\ \lambda &\geq 0. \end{aligned}$$

Let the optimal solution be c_o^* . Then (c_o^*, y_{1o}, y_{2o}) produces at least (y_{1o}, y_{2o}) at minimum cost $c_o^* = C(y_{1o}, y_{2o})$.

Now we check the economies of scope for this cost-efficient (c_o^*, y_o) as follows.

First, we create a set of virtual (artificial) diversified firms V by combining two firms from the groups S_1 and S_2 , in order to make comparisons among the diversified and specialized firm possible. For this purpose, let us arbitrarily choose two firms with cost and outputs represented by $(v_{1k}, z_{1k}) \in S_1$ and $(v_{2h}, z_{2h}) \in S_2$. Then we form a virtual diversified firm that is assumed to

produce both Product 1 (z_{1k}) and Product 2 (z_{2h}) with cost $v = v_{1k} + v_{2h}$. The set V of these virtual firms consists of all combinations of firms in S_1 and S_2 , giving $r (= p \times q)$ DMUs. Let us identify V as follows:

$$V = \{(v_j, z_{1j}, z_{2j}) | j = 1, \dots, r (= p \times q)\}. \quad (14.7)$$

The production possibility set corresponding to V for the virtual diversified firms is defined as

$$P_V = \{(c, y_1, y_2) | c = \sum_{j=1}^r v_j \lambda_j, y \leq \sum_{j=1}^r z_j \lambda_j, e\lambda = 1, \lambda \geq 0\}, \quad (14.8)$$

where v_j is the scalar $v_{1j} + v_{2j}$ while z_j and y are vectors where $z_j = \begin{pmatrix} z_{1j} \\ z_{2j} \end{pmatrix}$

represents the output of the j^{th} virtual diversified firm and $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ represents the outputs, i.e., Products 1 and 2.

Thus, we have two sets of diversified firms as displayed in Figure 14.2.

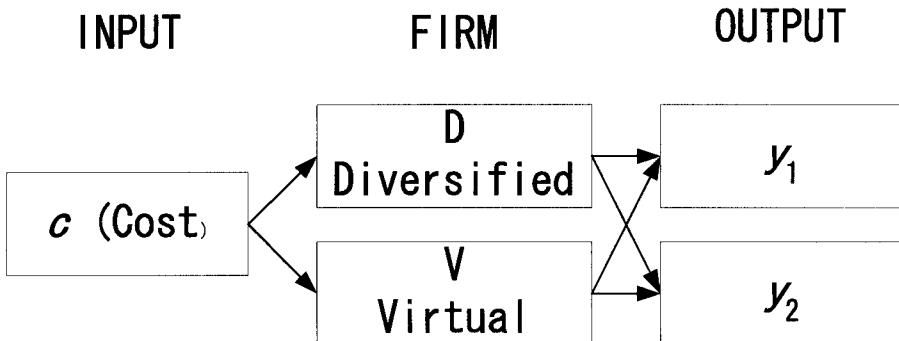


Figure 14.2. Diversified Firms D and Virtual Diversified Firms V

We evaluate the cost-efficiency of (c_o^*, y_{1o}, y_{2o}) with respect to sets of the virtual diversified firms in V as follows:

$$\begin{aligned} [\text{Cost-V}] \quad & \min_{\theta, \lambda} c_o^* \theta && (14.9) \\ \text{subject to} \quad & \theta c_o^* = v_1 \lambda_1 + \dots + v_r \lambda_r \\ & y_{1o} \leq z_{11} \lambda_1 + \dots + z_{1r} \lambda_r \\ & y_{2o} \leq z_{21} \lambda_1 + \dots + z_{2r} \lambda_r \\ & 1 = \lambda_1 + \dots + \lambda_r \\ & \lambda \geq 0, \end{aligned}$$

where c_o^* is obtained from (14.6).

Let the optimal θ be θ^* . Using this optimal θ^* , we define

$$C_1(y_{1o}, 0) + C_2(0, y_{2o}) = c_o^* \theta^*. \quad (14.10)$$

We have four cases. (i) $\theta^* < 1$. (c_o^*, y_{1o}, y_{2o}) is enveloped by P_V and hence is inefficient with respect to the virtual diversified group V. (ii) $\theta^* = 1$. (c_o^*, y_{1o}, y_{2o}) is on the radially efficient frontier of P_V . (iii) $\theta^* > 1$. (c_o^*, y_{1o}, y_{2o}) is not enveloped by V and hence is super-efficient with respect to V. (iv) The LP above has no feasible solution, i.e., either $y_{1o} > \max_{j=1}^r \{z_{1j}\}$ or $y_{2o} > \max_{j=1}^r \{z_{2j}\}$ so $e\lambda = 1$ cannot be satisfied. In this case, we assume $\theta^* = 1$ and hence $C_1(y_{1o}, 0) + C_2(0, y_{2o}) = c_{1o}^* = C(y_{1o}, y_{2o})$.

Using $C(y_{1o}, y_{2o})$ and $C_1(y_{1o}, 0) + C_2(0, y_{2o})$ we can check the local economies of scope at DMU_o by the conditions:

$$C(y_{1o}, y_{2o}) < C_1(y_{1o}, 0) + C_2(0, y_{2o}) \text{ Economies of scope exists.} \quad (14.11)$$

$$C(y_{1o}, y_{2o}) = C_1(y_{1o}, 0) + C_2(0, y_{2o}) \text{ Indifference.} \quad (14.12)$$

$$C(y_{1o}, y_{2o}) > C_1(y_{1o}, 0) + C_2(0, y_{2o}) \text{ Diseconomies of scope exists,} \quad (14.13)$$

with non-solvability included in the interpretation of (14.12).

In each case, DES (Degrees of Economies of Scope) is calculated by using (14.2) in the following way,

$$\text{DES}_j = \theta^* - 1. \quad (14.14)$$

Here we notice that we need not create *all* combinations of specialized firms in S₁ and S₂ as described above only a subset of them is sufficient to define P_V . This can be demonstrated as follows. Since P_1 is determined by a finite number of BCC-efficient DMUs in P_1 , an inefficient DMU $(v_{1\alpha}, z_{1\alpha})$ in P_1 can be expressed as a convex combination of the BCC-efficient DMUs with at least one non-zero slack in input or output as:

$$v_{1\alpha} = \sum_{j \in E_1} v_{1j} \lambda_j + s_\alpha^- \quad (14.15)$$

$$z_{1\alpha} = \sum_{j \in E_1} z_{1j} \lambda_j - s_\alpha^+ \quad (14.16)$$

$$1 = \sum_{j \in E_1} \lambda_j, \quad \lambda_j \geq 0 \ (\forall j) \quad (14.17)$$

$$(s_\alpha^-, s_\alpha^+) \neq (0, 0), \quad s_\alpha^- \geq 0, \quad s_\alpha^+ \geq 0, \quad (14.18)$$

where E_1 is the set of BCC-efficient DMUs in S₁. If we combine this inefficient DMU with $(v_{1\alpha}, z_{1\alpha})$ in P_1 with a DMU $(v_{1\beta}, z_{1\beta})$ in P_2 , we have a virtual DMU as follows:

$$(v_{1\alpha} + v_\beta, z_{1\alpha}, z_{2\beta}) = \sum_{j \in E_1} (v_{1j} + v_{2\beta}, z_{1j}, z_{2\beta}) \lambda_j + (s_\alpha^-, -s_\alpha^+, 0).$$

Since $(v_{1j} + v_{2\beta}, z_{1j}, z_{2\beta})$ is the combination of (v_{1j}, z_{1j}) with $(v_{2\beta}, z_{2\beta})$, it belongs to P_V and hence its convex combination, the first term on the right, is also included in P_V . By the assumption $(s_\alpha^-, s_\alpha^+) \neq (0, 0)$, the virtual DMU $(v_{1\alpha} + v_{2\beta}, z_{1\alpha}, z_{2\beta})$ has non-zero slack in the convex combined DMU and hence is not efficient with respect to P_V . Thus, we need not combine an inefficient DMU in S_1 with DMUs in S_2 .

Likewise, inefficient DMUs in S_2 do not contribute to the formation of P_V . Thus, we have:

Theorem 14.1 (Tone (2005)) *The virtual production possibility set P_V as defined in (14.8) can be formed from the combination of only the BCC-efficient DMUs in S_1 and S_2 .*

This can considerably reduce the number of virtual firms to be generated as well as the amount of computation for deciding whether economies of scope are present.

There is a possibility that the efficient frontiers of the generated virtual diversified firms fluctuate sharply because of outliers in the specialized firms so the decision on local economies of scope is unstable. Grosskopf and Yaisawarng (1990) proposed a scheme in response to this criticism. Their response takes the following form: (1) generate as many virtual diversified firms as possible from the groups S_1 and S_2 , e.g., pairwise additions and subtractions within each group and inter group to construct all possible combinations, and (2) using bootstrapping, randomly choose a fixed number of DMUs from the thus generated virtual firms and (3) use bootstrapping to repeat this process many times. In this way, they can obtain sampling distributions of degrees of economies of scope from which the mean scope values and their significance levels can be estimated. (Although we discussed the economies of scope in the two outputs case, we can extend this scheme to three or more outputs by generating virtual firms with three or more outputs from specialized ones in the same way as above.)

14.2.3 Checking a Virtual Merger

We can apply the above procedure in reverse to check the profitability of a merger. Suppose that we are considering whether to merge a Product 1-specialized firm (v_{1k}, z_{1k}) with a Product 2-specialized firm (v_{2h}, z_{2h}) .

First we remove cost-inefficiency from the two firms to obtain cost-efficient values $(v_{1k}^*, z_{1k}) \in P_1$ and $(v_{2h}^*, z_{2h}) \in P_2$ by applying the input-oriented BCC model for this single input (cost) and single output problem. Let $C_1(z_1, 0)$ ($C_2(0, z_2)$) be the minimum cost for producing z_{1k} (z_{2h}) in P_1 (P_2).

Then, we form a virtual firm consisting of one input (cost) ($v = v_{1k}^* + v_{2h}^*$) and two outputs ($z_1 = z_{1k}$, $z_2 = z_{2h}$). We check the economies of scope of this virtual firm in the set $P_{1,2}$ by solving the following cost model:

$$\begin{aligned} [\text{Cost-Merger}] \quad C(z_1, z_2) &= \min_{\theta, \lambda} v\theta \\ &\quad c_1\lambda_1 + c_2\lambda_2 + \cdots + c_n\lambda_n \end{aligned} \tag{14.19}$$

$$\begin{aligned} z_1 &\leq y_{11}\lambda_1 + y_{12}\lambda_2 + \cdots + y_{1n}\lambda_n \\ z_2 &\leq y_{21}\lambda_1 + y_{22}\lambda_2 + \cdots + y_{2n}\lambda_n \\ 1 &= \lambda_1 + \lambda_2 + \cdots + \lambda_n \\ \boldsymbol{\lambda} &\geq \mathbf{0}. \end{aligned}$$

In case of infeasibility, we set $C_1(z_1, 0) + C_2(0, z_2) = C(z_1, z_2)$.

Using $C(z_1, z_2)$, $C_1(z_1, 0)$ and $C_2(0, z_2)$ we can check the desirability of the merger by the conditions:

$$C(z_1, z_2) > C_1(z_1, 0) + C_2(0, z_2) \text{ Merger is unfavorable.} \quad (14.20)$$

$$C(z_1, z_2) = C_1(z_1, 0) + C_2(0, z_2) \text{ Indifference.} \quad (14.21)$$

$$C(z_1, z_2) < C_1(z_1, 0) + C_2(0, z_2) \text{ Merger is favorable.} \quad (14.22)$$

In each case, DES is calculated by (14.2).

14.2.4 Comparisons of Business Models

The aforementioned model deals with a single input (cost) and two outputs which can be produced by a diversified firm group or by two specialized firm groups. There is another way of approaching the use of this model. Suppose that a set of firms utilizes common inputs to produce common outputs. For example, hospitals use doctors, nurses, beds, technical staff, etc., as inputs, and process outpatients, inpatients, etc., as outputs. However, they have their scope of business via special characteristics. Some hospitals are classified into medicine and surgery, others into medicine and gynecology, and so forth. Thus, they have different business models.

Let us denote a subset of firms having a special scope of business by S and the rest by T. This scheme can be diagrammed as in Figure 14.3 in the case of two inputs and three outputs. We can evaluate these two distinct type of business as though they utilize the same inputs and produce the same outputs.

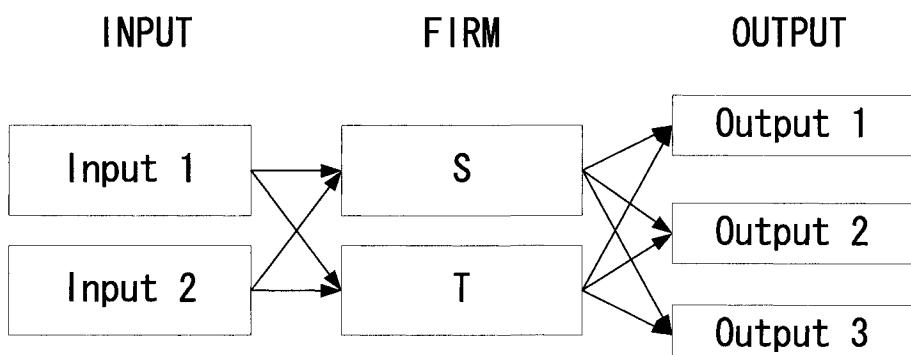


Figure 14.3. Firms with Two Different Scopes

Now, we compare the efficiency of S and T using the Bilateral Comparisons models based on the non-parametric Wilcoxon-Mann-Whitney rank-sum-test described in Section 7.6 of Chapter 7 as follows.

1. Eliminate technical inefficiencies from S and T using an appropriate DEA model and obtain a set of technically efficient DMUs on the efficient frontiers of each group.
2. Compare the thus obtained two sets of frontier efficient DMUs using the Bilateral Comparisons model.

If we skip Step 1 in the above scheme, we can evaluate the differences by reference to average efficiencies.

14.2.5 Illustrative Examples

Economies of Scope of Electric Utility Firms

Table 14.1 exhibits the performances of three groups belonging to electric utility firms. The first and the second groups are specialized respectively to (1) generation and (2) distribution of electric power, while the third group is diversified and handles both businesses.

(1) Identifying Economies of Scope of Group 3 Firms

In order to identify economies of scope of the diversified firms in Group 3, we first generate virtual firms by combining Group 1 and Group 2 firms. For this purpose, we applied the input-oriented BCC model to each group and found that Group 1 has four BCC-efficient firms — AEP Generating, Southern Electric Generating, South Carolina Generating and Safe Harbor Water Power —, while Group 2 has five efficient firms; PECO Energy, Baltimore Gas & Electric, Central Maine Power, Commonwealth Electric and Orange & Rockland Utilities.

From this result, we created $p \times q = 4 \times 5 = 20$ virtual firms as displayed in Table 14.2. For example, V1 is created by combining AEP Generating and PECO Energy.

Next, we eliminated inefficiency from diversified Group 3 using the input-oriented BCC model, and applied the [Cost-V] model in (14.9) for the BCC efficient Group 3 firms. Table 14.3 displays the results in the form of (1) efficiency measure θ^* , (2) DES (degree of economies of scope) and (3) we designate whether economies of scope are present. In this case [Cost-V] was feasible for every DMU. See the value of θ^* listed in the second column. It is observed that six firms display local economies of scope, while other two firms display diseconomies of scope. As a whole, the average DES is 0.138 (13.8%) showing existence of economies of scope in this illustrative example.

(2) Checking Virtual Mergers

To check these results, we evaluate the merger of South Carolina Generat-

Table 14.1. Data for Groups 1 and 2 as Specialized and Group 3 as Diversified Firms

Company Name	Group 1 Generation	
	Cost Prod ^a	Generation ^b
AEP Generating	253	832
Southern Electric Generating	179	682
South Carolina Generating	104	443
North Atlantic Energy	103	273
Ocean State Power	74	120
Ocean State Power II	69	124
Holyoke Water Power	35	98
Safe Harbor Water Power	30	97
Company Name	Group 2 Distribution	
	Cost Dist ^a	Customers served ^c
PECO Energy	951	1,529
Potomac Electric Power	549	718
Baltimore Gas & Electric	456	1,162
Boston Edison	412	687
Duquesne Light	203	586
Central Maine Power	159	560
Commonwealth Electric	95	355
Orange & Rockland Utilities	90	211
	Group 3 Generation and Distribution	
	Total Cost ^a	Generation ^b
Allele Inc.	277	771
Atlantic City Electric	246	250
Central Vermont Public Service	79	38
El Paso Electric	314	779
Kentucky Power	183	575
New York State Electric & Gas	335	26
NorthWestern Energy	104	114
Ohio Edison	574	820

^a All cost units are stated in one million US\$. ^b Generation unit: MWh \times 10000. ^c Customers unit: 10000 households.

ing (Generation) with Commonwealth Electric (Distribution), with both being efficient in the groups they belong to. This merged firm shows costs of $c = 104 + 95 = 199$ to produce $z_1 = 443$ (Generation) and $z_2 = 355$ (Customers). See Groups 1 and 2 in Table 14.1. Hence $C_1(443, 0) + C_2(0, 355) = 199$. We evaluated this data set using the model [Cost-Merger] and found that

$$C(443, 355) = 194 < 199 = C_1(443, 0) + C_2(0, 355).$$

Table 14.2. Twenty Virtual Diversified Firms

Firm	Cost	Generation	Customers	Firm	Cost	Generation	Customers
V1	1,204	832	1,529	V11	1,055	443	1,529
V2	709	832	1,162	V12	560	443	1,162
V3	412	832	560	V13	263	443	560
V4	348	832	355	V14	199	443	355
V5	344	832	211	V15	195	443	211
V6	1,130	682	1,529	V16	980	97	1,529
V7	635	682	1,162	V17	485	97	1,162
V8	338	682	560	V18	189	97	560
V9	274	682	355	V19	125	97	355
V10	270	682	211	V20	120	97	211

Table 14.3. Efficiency Score and Degree of Economies of Scope

Company	θ^*	DES	Economies of Scope
Alltec	1.134	0.134	Yes
Atlantic City Electric	0.896	-0.104	No
Central Vermont Public Service	1.514	0.514	Yes
El Paso Electric	1.022	0.022	Yes
Kentucky Power	1.290	0.290	Yes
New York State Electric & Gas	0.932	-0.068	No
NorthWestern Energy	1.223	0.223	Yes
Ohio Edison	1.094	0.094	Yes

This means that the diversified group can produce the same outputs at a lower cost than the merged firm and hence merger is not recommended, in this case.

14.3 CAPACITY UTILIZATION

In this section we introduce the concept of capacity utilization and discuss its technical and price-based aspects.

14.3.1 Fixed vs. Variable Input Resources

Among input resources for production, some are fixed and cannot be easily altered, e.g., invested capital, while some are flexibly alterable, e.g., employee or office hours. We call the former *fixed* and the latter *variable*.

We denote the output vector of DMU_j ($j = 1, \dots, n$) by $\mathbf{y}_j = (y_{1j}, \dots, y_{sj}) \in R^s$ and its fixed and variable input vectors by $\mathbf{x}_j^F = (x_{1j}^F, \dots, x_{fj}^F) \in R^f$ and $\mathbf{x}_j^V = (x_{1j}^V, \dots, x_{vj}^V) \in R^v$, respectively.

14.3.2 Technical Capacity Utilization

Given the set of observed DMUs $\{(\mathbf{x}_j^F, \mathbf{x}_j^V, \mathbf{y}_j)\}$ ($j = 1, \dots, n$), we evaluate the efficiency of DMU_o = $(\mathbf{x}_o^F, \mathbf{x}_o^V, \mathbf{y}_o)$ ($o = 1, \dots, n$) using the output-oriented non-radial model (SBM-O) described in (4.79) as follows.

[SBM_o-Restricted]

$$\phi^* = \max_{\lambda, s^+} \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}} \right) \quad (14.23)$$

subject to

$$x_{io}^F \geq x_{i1}^F \lambda_1 + \dots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.24)$$

$$x_{io}^V \geq x_{i1}^V \lambda_1 + \dots + x_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.25)$$

$$y_{ro} = y_{r1} \lambda_1 + \dots + y_{rn} \lambda_n - s_r^+ \quad (r = 1, \dots, s) \quad (14.26)$$

$$1 = \lambda_1 + \dots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j), \quad s_r^+ \geq 0 \quad (\forall r). \quad (14.27)$$

Let the optimal solution of [SBM_o-Restricted] be $(\lambda_1^*, \dots, \lambda_n^*)$ and $(s_1^{+*}, \dots, s_s^{+*})$, and define:

$$\tilde{x}_{io}^F = x_{i1}^F \lambda_1^* + \dots + x_{in}^F \lambda_n^* \quad (i = 1, \dots, f) \quad (14.28)$$

$$\tilde{x}_{io}^V = x_{i1}^V \lambda_1^* + \dots + x_{in}^V \lambda_n^* \quad (i = 1, \dots, v) \quad (14.29)$$

$$\tilde{y}_{ro} = y_{r1} \lambda_1^* + \dots + y_{rn} \lambda_n^*. \quad (r = 1, \dots, s) \quad (14.30)$$

These values $(\tilde{\mathbf{x}}_o^F, \tilde{\mathbf{x}}_o^V, \tilde{\mathbf{y}}_o)$ represent the coordinates of a point on the efficient frontier of the production possibility set defined by

$$P = \{(\mathbf{x}^F, \mathbf{x}^V, \mathbf{y})\}, \quad (14.31)$$

with

$$x_i^F \geq x_{i1}^F \lambda_1 + \dots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.32)$$

$$x_i^V \geq x_{i1}^V \lambda_1 + \dots + x_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.33)$$

$$y_r \leq y_{r1} \lambda_1 + \dots + y_{rn} \lambda_n \quad (r = 1, \dots, s) \quad (14.34)$$

$$1 = \lambda_1 + \dots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j). \quad (14.35)$$

Let us define ϕ_r^* by

$$\phi_r^* = 1 + \frac{s_{ro}^*}{y_{ro}} (\geq 1) \quad (r = 1, \dots, s). \quad (14.36)$$

Then

$$\phi^* = \frac{1}{s} \sum_{r=1}^s \phi_r^*. \quad (14.37)$$

This means that the optimal value ϕ^* is the average expansion rate ϕ_r^* ($r = 1, \dots, s$) of the observed output y_o . Here $\phi^* > 1$ means that the DMU_{*o*} can expand some outputs without worsening other outputs, using the current level of input resources.

We now relax the model (14.23)-(14.27) by deleting the constraint (14.25), which bounds the variable input values below the observed values. This yields the following model.

[SBM_{*o*}-Relaxed]

$$\phi_F^* = \max_{\lambda, s^+} \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{y_{ro}} \right) \quad (14.38)$$

subject to

$$x_{io}^F \geq x_{i1}^F \lambda_1 + \dots + x_{in}^F \lambda_n \quad (i = 1, \dots, v) \quad (14.39)$$

$$y_{ro} = y_{r1} \lambda_1 + \dots + y_{rn} \lambda_n - s_r^+ \quad (r = 1, \dots, s)$$

$$1 = \lambda_1 + \dots + \lambda_n, \quad \lambda_j \geq 0 \ (\forall j), \quad s_r^+ \geq 0 \ (\forall r).$$

Let the optimal solution of [SBM_{*o*}-Relaxed] be $(\lambda_1^{**}, \dots, \lambda_n^{**})$ and $(s_1^{+**}, \dots, s_s^{+**})$. For this solution, the amount of the variable inputs is given by

$$\hat{x}_{io}^V = x_{i1}^V \lambda_1^{**} + \dots + x_{in}^V \lambda_n^{**}, \quad (i = 1, \dots, v) \quad (14.40)$$

which may be greater or less than the observed input value x_{io}^V recorded in (14.25).

Let us define a capacity utilization measure by

$$\kappa_o^* = \frac{\phi^*}{\phi_F^*} (\leq 1), \quad (14.41)$$

where ϕ^* is exhibited in (14.23) and ϕ_F^* is fixed in (14.38).

Then, $\kappa_o^* < 1$, i.e., $\phi_F^* > \phi^*$ indicates that some of the bounding constraints (14.25) are tight in the model [SBM_{*o*}-Restricted] and the average expansion rate can be increased by deleting these bounds for the variable inputs. We further define a measure of capacity utilization for each variable input *i* as

$$\delta_{io}^* = \frac{\hat{x}_{io}^V}{x_{io}^V} \quad (14.42)$$

Under the constant returns-to-scale assumption, Färe, Grosskopf and Lovell (1994) introduced the above scheme using the radial (CCR-O) model and named this ratio κ_o^* the *Plant Capacity Utilization Measure* of the DMU_{*o*} that was evaluated. Furthermore, they called the ratio δ_{io}^* the *ith Variable Input Utilization Rate*.

14.3.3 Price-Based Capacity Utilization Measure

We now extend this analysis to the case when the factor prices of inputs and outputs are available. Let the prices of the output y_{ro} and the variable input

x_{io}^V be p_{ro} and c_{io}^V , respectively, which are not identical among DMUs. Then, the current profit of DMU_o regarding the revenue from the observed outputs and the cost for the variable inputs can be expressed as

$$G_o = \sum_{r=1}^s p_{ro} y_{ro} - \sum_{i=1}^v c_{io}^V x_{io}^V. \quad (o = 1, \dots, n) \quad (14.43)$$

In order to gauge the maximum profit attainable within the price-based production possibility set, we reformulate our problem in line with the New-Profit model described in Section 8.3.4 in Chapter 8.

First, we define a new production possibility set \bar{P} as

$$\bar{P} = \{(\mathbf{x}^F, \bar{\mathbf{x}}^V, \bar{\mathbf{y}})\}, \quad (14.44)$$

with

$$x_i^F \geq x_{i1}^F \lambda_1 + \dots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.45)$$

$$\bar{x}_i^V \geq \bar{x}_{i1}^V \lambda_1 + \dots + \bar{x}_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.46)$$

$$\bar{y}_r \leq \bar{y}_{r1} \lambda_1 + \dots + \bar{y}_{rn} \lambda_n \quad (r = 1, \dots, s) \quad (14.47)$$

$$1 = \lambda_1 + \dots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j),$$

where \bar{x}_{ij}^V and \bar{y}_{rj} are respectively defined by

$$\bar{x}_{ij}^V = c_{ij}^V x_{ij}^V \quad (i = 1, \dots, v : j = 1, \dots, n) \quad (14.48)$$

$$\bar{y}_{rj} = p_{rj} y_{rj} \quad (r = 1, \dots, s : j = 1, \dots, n). \quad (14.49)$$

We solve the following three LPs for each DMU_o ($o = 1, \dots, n$) and therefore solve $3n$ LPs in total,

[New-Tech_o]

$$\phi^* = \max_{\lambda, \bar{s}^+} \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{\bar{s}_r^+}{\bar{y}_{ro}} \right) \quad (14.50)$$

subject to

$$x_{io}^F \geq x_{i1}^F \lambda_1 + \dots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.51)$$

$$\bar{x}_{io}^V \geq \bar{x}_{i1}^V \lambda_1 + \dots + \bar{x}_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.52)$$

$$\bar{y}_{ro} = \bar{y}_{r1} \lambda_1 + \dots + \bar{y}_{rn} \lambda_n - \bar{s}_r^+ \quad (r = 1, \dots, s) \quad (14.53)$$

$$1 = \lambda_1 + \dots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j), \quad \bar{s}_r^+ \geq 0 \quad (\forall r). \quad (14.54)$$

Using the optimal solution $(\bar{\lambda}_1, \dots, \bar{\lambda}_n)$ of [New-Tech_o], the efficient $(\bar{\bar{x}}_o^F, \bar{\bar{x}}_o^V, \bar{\bar{y}}_o)$ is defined by

$$\bar{\bar{x}}_{io}^F = x_{i1}^F \bar{\lambda}_1 + \dots + x_{in}^F \bar{\lambda}_n \quad (i = 1, \dots, f) \quad (14.55)$$

$$\bar{\bar{x}}_{io}^V = \bar{x}_{i1}^V \bar{\lambda}_1 + \dots + \bar{x}_{in}^V \bar{\lambda}_n \quad (i = 1, \dots, v) \quad (14.56)$$

$$\bar{\bar{y}}_{ro} = \bar{y}_{r1} \bar{\lambda}_1 + \dots + \bar{y}_{rn} \bar{\lambda}_n. \quad (r = 1, \dots, s). \quad (14.57)$$

Since

$$\bar{x}_{io}^V \leq \bar{x}_{io}^V = c_{io}^V x_{io}^V \text{ and } \bar{y}_{ro} \geq \bar{y}_{ro} = p_{ro} y_{ro}, \quad (14.58)$$

we have the following profit G_o^* and inequality relations,

$$G_o^* = \sum_{r=1}^s \bar{y}_{ro} - \sum_{i=1}^v \bar{x}_{io}^V \geq \sum_{r=1}^s p_{ro} y_{ro} - \sum_{i=1}^v c_{io}^V x_{io}^V = G_o. \quad (14.59)$$

Now we can define the loss of profit L_o due to technical inefficiency in the production possibility set \bar{P} as follows:

$$L_o = G_o^* - G_o (\geq 0). \quad (14.60)$$

Secondly, we solve the New-Profit model described in Section 8.3.4 on the \bar{P} .

[Profit_o-Restricted]

$$G_o^{**} = \max_{\lambda, \bar{x}^V, \bar{y}} \sum_{r=1}^s \bar{y}_r - \sum_{i=1}^v \bar{x}_i^V \quad (14.61)$$

subject to

$$x_{io}^F \geq x_{i1}^F \lambda_1 + \cdots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.62)$$

$$\bar{x}_i^V = \bar{x}_{i1}^V \lambda_1 + \cdots + \bar{x}_{in}^V \lambda_n \leq \bar{x}_{io}^V \quad (i = 1, \dots, v) \quad (14.63)$$

$$\bar{y}_r = \bar{y}_{r1} \lambda_1 + \cdots + \bar{y}_{rn} \lambda_n \geq \bar{y}_{ro} \quad (r = 1, \dots, s) \quad (14.64)$$

$$1 = \lambda_1 + \cdots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j).$$

We define

$$L_o^* = G_o^{**} - G_o^* (\geq 0). \quad (14.65)$$

L_o^* is the loss of profit due to the allocative (price-based input-output mix) inefficiency.

Lastly, we relax the constraint (14.63) and form [Profit_o-Relaxed] as follows.

[Profit_o-Relaxed]

$$G_o^{***} = \max_{\lambda, \bar{x}^V, \bar{y}} \sum_{r=1}^s \bar{y}_r - \sum_{i=1}^v \bar{x}_i^V \quad (14.66)$$

subject to

$$x_{io}^F \geq x_{i1}^F \lambda_1 + \cdots + x_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.67)$$

$$\bar{x}_i^V = \bar{x}_{i1}^V \lambda_1 + \cdots + \bar{x}_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.68)$$

$$\bar{y}_r = \bar{y}_{r1} \lambda_1 + \cdots + \bar{y}_{rn} \lambda_n \geq \bar{y}_{ro} \quad (r = 1, \dots, s) \quad (14.69)$$

$$1 = \lambda_1 + \cdots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j).$$

We define

$$L_o^{**} = G_o^{***} - G_o^* (\geq 0). \quad (14.70)$$

L_o^{**} is the loss of profit due to the price-based capacity utilization inefficiency.

Thus we find the global maximum profit G_o^{***} for the DMU_{*o*} and its decomposition as follows:

$$G_o^{***} = G_o^{**} + L_o^* \quad (L_o^* = \text{Loss due to capacity utilization}) \quad (14.71)$$

$$= G_o^* + L_o^* + L_o^* \quad (L_o^* = \text{Loss due to allocative inefficiency}) \quad (14.72)$$

$$= G_o + L_o^* + L_o^* + L_o \quad (L_o = \text{Loss due to technical inefficiency}).$$

$$(14.73)$$

Figure 14.4 illustrates this model with a single variable input \bar{x}^V and a single output \bar{y} . The observed point A = (\bar{x}_o^V, \bar{y}_o) has the profit G_o as displayed by the dotted line passing through A (an iso-profit line). This point moves up to B which is the optimum solution of both [New-Tech_{*o*}] and [Profit_{*o*}-Restricted] with the profit $G_o^* = G_o^{**}$ in this case. Lastly, by relaxing the constraint on \bar{x}^V , we move B up to C along the efficient frontier with the maximum profit G_o^{***} .

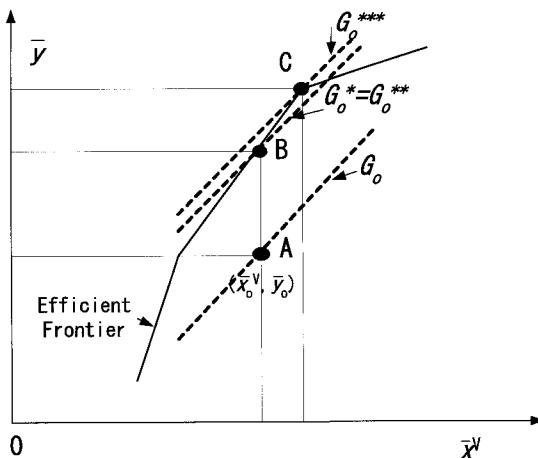


Figure 14.4. Price-Based Capacity Utilization

Although we take no account of the unit-price of the fixed inputs, we can modify the above models if these prices are available, as follows. Let us define $\bar{P} = \{(\bar{x}^F, \bar{x}^V, \bar{y})\}$ with $\bar{x}_{ij}^F = c_{ij}^F x_{ij}^F$ where c_{ij}^F is the unit price of the fixed variable x_{ij}^F . We replace x_{ij}^F by \bar{x}_{ij}^F in the above scheme. This modification will result in changes in the optimal solutions which are more price-oriented than the former scheme.

14.3.4 Long-Run and Short-Run Capacity Utilization

Note that the objective function of model [Profit_{*o*}-Relaxed] represents profit in the *short-run* where the fixed input factor like capital is kept constant. One can, however, relax this assumption of *fixity* on capital input in the long run where all the inputs are variable. The long-run profit can be obtained from the

following:

[Profit_o-Long-Run]

$$G_o^{****} = \max_{\lambda, \bar{x}^F, \bar{x}^V, \bar{y}} \sum_{r=1}^s \bar{y}_r - \sum_{i=1}^f \bar{x}_i^F - \sum_{i=1}^v \bar{x}_i^V \quad (14.74)$$

subject to

$$\bar{x}_i^F = \bar{x}_{i1}^F \lambda_1 + \cdots + \bar{x}_{in}^F \lambda_n \quad (i = 1, \dots, f) \quad (14.75)$$

$$\bar{x}_i^V = \bar{x}_{i1}^V \lambda_1 + \cdots + \bar{x}_{in}^V \lambda_n \quad (i = 1, \dots, v) \quad (14.76)$$

$$\bar{y}_r = \bar{y}_{r1} \lambda_1 + \cdots + \bar{y}_{rn} \lambda_n \geq \bar{y}_{ro} \quad (r = 1, \dots, s) \quad (14.77)$$

$$1 = \lambda_1 + \cdots + \lambda_n, \quad \lambda_j \geq 0 \quad (\forall j).$$

We now define

$$L_o^{***} = G_o^{****} - G_o^{***} (\geq 0). \quad (14.78)$$

L_o^{***} is the loss of profit due to the gap in short and long runs.

Thus, we find the global maximum profit G_o^{****} for the DMU_o and the gap of actual profit (G_o) from global maximum profit G_o^{****} can be decomposed as follows:

$$G_o^{****} = G_o^{***} + L_o^{***} \quad (L_o^{***} = \text{Loss due to short-long run gap}) \quad (14.79)$$

$$= G_o^{**} + L_o^{***} + L_o^{**} \quad (L_o^{**} = \text{Loss due to capacity utilization}) \quad (14.80)$$

$$= G_o^* + L_o^{***} + L_o^{**} + L_o^* \quad (L_o^* = \text{Loss due to allocative inefficiency}). \quad (14.81)$$

$$= G_o + L_o^{***} + L_o^{**} + L_o^* + L_o \quad (L_o = \text{Loss due to technical inefficiency}). \quad (14.82)$$

14.3.5 Illustrative Examples

Table 14.4 exhibits a data set of twelve hospitals with one fixed input *Doctor*, one variable input *Nurse*, and two outputs; *Outpatient* and *Inpatient*, along with their unit prices. Since *Doctor* is assumed to be fixed, we did not record their unit prices.

1. Measure of Technical Capacity Utilization

We now illustrate the measure of “technical utilization” using a hospital data set in which we assume the number of *Doctors* is fixed, (i.e., the upper bound of doctors for each hospital is fixed) and the number of *Nurses* is variable. Hence we are evaluating the capacity utilization of nurses. The output-oriented non-radial model under VRS; [SBM_o-Restricted] (14.23) was applied to measure the technical efficiency of each hospital. In this case, we made no distinction between fixed and variable inputs. As usual they dealt with inputs. The ϕ^* scores in Table 14.5 exhibit the results. $\phi^* = 1$ indicates technical efficient DMUs, while $\phi^* > 1$ indicates inefficient DMUs. Next we deleted the constraint on the number of nurses and solved the model [SBM_o-Relaxed] (14.38). The

Table 14.4. Data for 12 Hospitals

DMU	Inputs			Outputs			
	Doctor(Fixed) Number	Nurse(Variable) Number	Price	Outpatient Number	Price	Inpatient Number	Price
A	20	151	100	100	550	90	2010
B	19	131	80	150	400	50	1800
C	25	160	90	160	480	55	2200
D	27	168	120	180	600	72	3500
E	22	158	70	94	400	66	3050
F	55	255	80	230	430	90	3900
G	33	235	100	220	540	88	3300
H	31	206	85	152	420	80	3500
I	30	244	76	190	350	100	2900
J	50	268	75	250	410	100	2600
K	53	306	80	260	540	147	2450
L	38	284	70	250	295	120	3000

score ϕ_F^* in the table exhibits the results. Comparisons between columns 1 and 2 in Table 14.5 shows that hospitals C, D, F, G, H and J increased (worsened) the score and thereby demonstrated that these hospitals can increase the number of outpatients and/or inpatients if they increase the number of nurses. With the ratio ϕ^*/ϕ_F^* , we obtained the measure of capacity utilization κ^* in the table. As in (14.41) $\kappa^* < 1$ indicates under-utilization (shortfalls) of nurses. Lastly, we computed the nurse utilization rate δ^* . As defined in (14.42) $\delta^* > 1$ and $\delta^* < 1$ respectively indicate shortfalls and excesses of nurses.

2. Price-Based Measure of Capacity Utilization

Using the price information in Table 14.4, we form the price-based data set $\{(x_j^F, \bar{x}_j = c_j^V x_j^V, \bar{y}_{1j} = p_{1j} y_{1j}, \bar{y}_{2j} = p_{2j} y_{2j})\}$ that is displayed in Table 14.6. Since *Doctor* is assumed to be fixed in this example, we do not account for their cost. The current profit is calculated as $G = \bar{y}_1 + \bar{y}_2 - \bar{x}^V$ for each DMU. Again we notice in this illustrative example that interest centers on the capacity utilization of nurses for each hospital, i.e., the amount of increase in the expected profit induced by relaxing the bound of nurse utilization.

We solved three LPs [New-Tech], [Profit-Restricted] and [Profit-Relaxed] for each DMU and obtained the corresponding profits G^* (technically efficient profit), G^{**} (allocative efficient profit) and G^{***} (capacity relaxed maximum profit) as well as losses L (loss due to technical inefficiency), L^* (loss due to allocative inefficiency) and L^{**} (loss due to capacity utilization), that are displayed in Table 14.7.

Table 14.5. Technical Capacity Utilization

DMU	Restricted ϕ^*	Relaxed ϕ_F^*	Capacity Utilization $\kappa^* (= \phi^*/\phi_F^*)$	Nurse Utilization Rate (δ^*)
A	1.000	1.000	1.000	1.000
B	1.000	1.000	1.000	1.000
C	1.163	1.284	0.906	1.150
D	1.000	1.107	0.903	1.179
E	1.342	1.342	1.000	0.992
F	1.164	1.382	0.842	1.200
G	1.000	1.091	0.917	1.041
H	1.219	1.308	0.932	1.128
I	1.011	1.011	1.000	0.916
J	1.000	1.224	0.817	1.125
K	1.000	1.000	1.000	1.000
L	1.000	1.000	1.000	1.000

Table 14.6. Price-Based Data Set

DMU	Doctor x^F	Nurse \bar{x}^V	Outpat \bar{y}_1	Inpat \bar{y}_2
A	20	15,100	55,000	180,900
B	19	10,480	60,000	90,000
C	25	14,400	76,800	121,000
D	27	20,160	108,000	252,000
E	22	11,060	37,600	201,300
F	55	20,400	98,900	351,000
G	33	23,500	118,800	290,400
H	31	17,510	63,840	280,000
I	30	18,544	66,500	290,000
J	50	20,100	102,500	260,000
K	53	24,480	140,400	360,150
L	38	19,880	73,750	360,000

The results indicate that hospitals C, E, F, H and J can improve their profits by relaxing the capacity (nurse) constraint and the losses due to inefficient capacity utilization is larger than the other losses in this example.

Furthermore, Table 14.8 displays the current and maximum cost, revenue and profit for each DMU. These figures indicate how the increase of the nurse cost can contribute to the growth of the outpatient and inpatient revenues.

Table 14.7. Profits and Losses

	Profit				Loss				Total
	Current G	Tech G*	Restrict G**	Relaxed G***	Tech L	Alloc L*	Capacity L**		
A	220,800	220,800	220,800	220,800	0	0	0	0	0
B	139,520	139,520	139,520	139,520	0	0	0	0	0
C	183,400	237,098	237,098	305,830	53,698	0	68,732	122,430	
D	339,840	339,840	339,840	339,840	0	0	0	0	0
E	227,840	227,840	227,840	254,212	0	0	26,372	26,372	
F	429,500	429,500	429,500	476,070	0	0	46,570	46,570	
G	385,700	385,700	385,700	385,700	0	0	0	0	0
H	326,330	339,193	345,123	369,772	12,863	5,931	24,649	43,442	
I	337,956	337,956	337,956	337,956	0	0	0	0	0
J	342,400	407,318	407,318	463,630	64,918	0	56,312	121,230	
K	476,070	476,070	476,070	476,070	0	0	0	0	0
L	413,870	413,870	413,870	413,870	0	0	0	0	0
					Sum	131,478	5,931	222,635	360,044

Table 14.8. Comparisons of Current and Maximum Profits

DMU	Current			Maximum			Loss
	Cost Nurse	Revenue Outpat	Inpat	Cost Nurse	Revenue Outpat	Inpat	
A	15,100	55,000	180,900	15,100	55,000	180,900	0
B	10,480	60,000	90,000	10,480	60,000	90,000	0
C	14,400	76,800	121,000	18,714	92,858	231,686	122,430
D	20,160	108,000	252,000	20,160	108,000	252,000	0
E	11,060	37,600	201,300	16,503	69,415	201,300	26,372
F	20,400	98,900	351,000	24,480	140,400	360,150	46,570
G	23,500	118,800	290,400	23,500	118,800	290,400	0
H	17,510	63,840	280,000	21,978	111,750	280,000	43,442
I	18,544	66,500	290,000	18,544	66,500	290,000	0
J	20,100	102,500	260,000	23,560	127,070	360,120	121,230
K	24,480	140,400	360,150	24,480	140,400	360,150	0
L	19,880	73,750	360,000	19,880	73,750	360,000	0

We explain the case of DMU H in detail which has the current profit regarding revenues from outpatients and inpatients and cost of nurses as

$$G_H = \text{Current profit} = 152 \times 420 + 80 \times 3,500 - 85 \times 206 = 326,330.$$

This is the profit obtained from the revenue and cost figures as $63,840 + 280,000 - 17,510 = 326,330$.

The optimal solution of [New-Tech_H] (14.50) for H is given by

$$\phi_H^* = 1.101, \lambda_D^* = 0.277, \lambda_E^* = 0.316, \lambda_K^* = 0.073, \lambda_L^* = 0.334.$$

Hence, the (price-based) technical efficient inputs and outputs are obtained using (14.55)-(14.57) as the optimal cost of nurses = $0.277 \times 20,160 + 0.316 \times 11,060 + 0.073 \times 24,480 + 0.334 \times 19,880 = 17,510$ which is equal to the current value. Similarly, we have the optimal number of doctors = 31 = current value. The optimal revenue from outpatients is 76,703 which is greater than the current value 63,840. The optimal revenue from inpatients is 280,000 which is equal to the current value. Thus, H can get a larger profit by deleting technical inefficiency in outpatients as

$$G_H^* = \text{Technically efficient profit} = 76,703 + 280,000 - 17,510 = 339,193.$$

(See the 3rd column of Table 14.7.)

Then we solve [Profit_H-Restricted] (14.61) for H in which cost of nurses is restricted by the current value 17,510 and obtained the optimal solution $G_H^{**} = 345,123$ with revenue from outpatients = 71,804 and revenue from inpatients = 290,829, and cost of nurses= 17,510. This solution indicates that relaxation of nurse capacity may give more profit to H.

Finally, we relaxed the nurse capacity (17,510) and solved [Profit_H-Relaxed] (14.66). The optimal solution is $G_H^{***} = 369,772$ with revenue from outpatients = 111,750 which is greater than the current value 71,804, revenue from inpatients = 280,000=the current value, and cost of nurses= 21,978 which is greater than the current cost 17,510. Thus, we can attain the maximum profit 369,772 by relaxing the cost-based nurse capacity from 17,510 to 21,978. (See Table 14.8.)

Figure 14.5 exhibits decomposition of the maximum profit G_o^{***} into the current profit G_o and three losses, i.e., losses due to technical (L_o), allocative (L_o^*) and capacity utilization (L_o^{**}) inefficiencies.

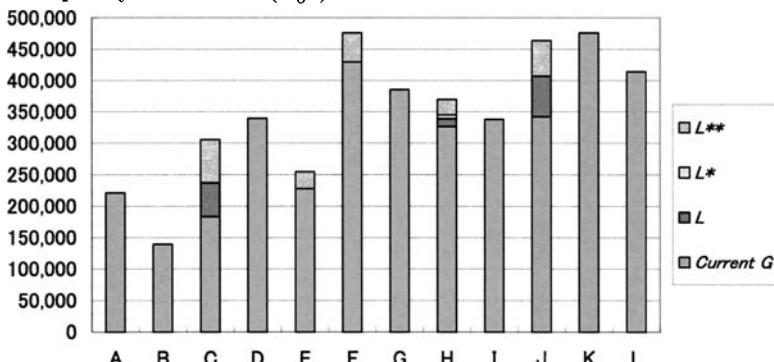


Figure 14.5. Decomposition of the Maximum Profit

14.4 SUMMARY OF CHAPTER 14

In this chapter, we have covered two subjects; economies of scope and capacity utilization.

1. Economies of scope concerns the problem whether any significant difference in efficiency exists between specialized and diversified firms. The specialized firms produce different (specialized) products, whereas the diversified firms produce all of the concerned products. We extended this model to comparisons of two business models where both models utilize common inputs for producing common outputs but their business styles are different.
2. Capacity utilization deals with the situation where some input values are fixed and cannot be easily altered while some are flexibly alterable. Thus, we divided inputs into fixed and variable inputs, and studied the effect on efficiency caused by the variable inputs according to whether relaxation of variable inputs improves the efficiency of DMU under evaluation. We analyzed two aspects of capacity utilization; (1) technical and (2) price-based.
3. Example applications were presented to illustrate uses of both these approaches.

14.5 NOTES AND SELECTED BIBLIOGRAPHY

Barua *et al.* (2004)⁶ apply DEA to determine relative efficiencies between internet dot com companies that produce only physical products and those that produce only digital products. To allow for the fact that the latter are relatively inexperienced, a distinction is made between long- and short-run efficiencies and inefficiencies, with a finding of no statistically significant difference in the short run but digital product companies are significantly more efficient in the long run. A new way of distinguishing between long- and short-run performances is thus utilized that avoids the need for identifying the time periods associated with long-run vs. short-run efficiencies and inefficiencies. In place of “time,” this paper utilizes differences in the “properties” that economic theory associates with long- and short-run performances.

Earlier discussions of capacity utilization are provided by Gold (1955)⁷ and Johansen (1968), and extended by several authors in the DEA framework, e.g., Färe *et al.* (1989a) (1989b)⁸ and Coelli *et al.* (2002).⁹ The SBM based capacity utilization analyses in Section 14.3 were developed by Tone (2005).¹⁰ Sahoo and Tone (2006)¹¹ reviews the existing methods of capacity utilization in a nonparametric framework from an economic perspective, and then suggests an alternative in the light of limitations therein. In the spirit of work by Coelli *et al.* (2002) they propose two methods, radial and non-radial, to decompose input-based technological capacity utilization into various meaningful components viz., technical inefficiency, ray economic capacity utilization and optimal capacity idleness. A case study of Indian banking industry is taken as an example to illustrate the potential application of these two methods of decomposition. Their two broad empirical findings are that first, competition

created after financial sector reforms generates high efficiency growth, and reduces excess capacity; second, the cost gap of short-run cost from actual cost is higher for nationalized banks over private banks indicating that the latter banks, though old, do not reflect their learning experience in their cost minimizing behavior.

14.6 RELATED DEA-SOLVER MODELS FOR CHAPTER 14

- Economies of Scope
 - The model [Cost] (14.6) can be solved by the input-oriented BCC model (BCC-I).
 - The models [Cost-W] (14.9) and [Cost-Merger] (14.19) can be solved by the input-oriented super BCC model (Super-BCC-I).
 - “Comparisons of Business Models” in Section 14.2.4 can be solved by the Bilateral-BCC-I code after eliminating the technical inefficiency using appropriate DEA models, e.g., BCC-I.
- Capacity Utilization
 - The models [SBM_o-Restricted] (14.23), [SBM_o-Relaxed] (14.38) and [New-Tech_o] (14.50) can be solved by the output-oriented SBM code under VRS (SBM-O-V).
 - Both [Profit_o-Restricted] (14.61) and [Profit_o-Relaxed] (14.66) can be solved by the new-profit model under VRS (New-Profit-V).

14.7 PROBLEM SUPPLEMENT FOR CHAPTER 14

Problem 14.1

In the capacity utilization model, we discriminated between fixed and variable factors only in inputs. Can you extend this to outputs in the price-based model?

Suggested Response : It is often observed that some outputs are fixed due to long term contracts but some are variable depending on market trends. Corresponding to these situations, we divide the output $y_j \in R^s$ into fixed and variable factors respectively as $y_j^f \in R^F$ and $y_j^v \in R^V$ ($j = 1, \dots, n$) with $F + V = s$. We denote the factor price of the variable output y_j^V by p_j^V ($j = 1, \dots, n$). We deal only with the variable inputs and outputs, and hence the current profit for the DMU_o is expressed as

$$G_o = \sum_{r=1}^V p_{ro}^V y_{ro}^V - \sum_{i=1}^v c_{io}^V x_{io}^V.$$

Correspondingly, we modify the production possibility set \overline{P} as

$$\overline{P} = \{(x^F, \bar{x}^V, y^F, \bar{y}^V)\}.$$

In the objective functions of [New-Tech_o], [Profit_o-Restricted] and [Profit_o-Relaxed], we employ only the \bar{y}_o^V related terms. Also, in the constraints parts, we decompose the \bar{y} related constraints into y^F and \bar{y}^V parts. Lastly, we delete the \bar{y}^V related constraints from [Profit_o-Relaxed]. Thus, the output constraint is replaced by

$$y_r^F = y_{r1}^F \lambda_1 + \cdots + y_{rn}^F \lambda_n \geq y_{ro}^F (r = 1, \dots, F).$$

Since we assume that the variable outputs can vary independently of the current values, it may be possible that we can increase profit by reducing some of the current variable outputs. Figure 14.6 illustrates such a situation in a single variable input and a single variable output case. We can arrive at the maximum profit G_o^{***} at the point C by reducing the variable output \bar{y}^V .

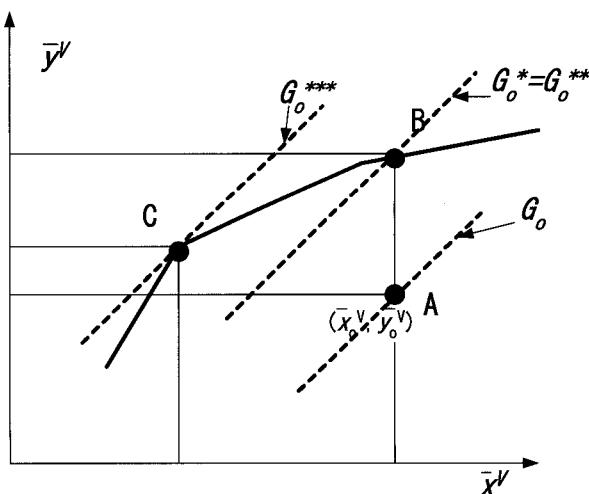


Figure 14.6. Price-Based Capacity Utilization with a Variable Output

Notes

1. W.J. Baumol, J.C. Panzar and R.D. Willig (1982), *Contestable Markets and the Theory of Industrial Structure*, New York: Harcourt Brace Jovanovich.
2. R. Färe (1986), "Addition and Efficiency," *Quarterly Journal of Economics* 101, pp.861-866, and Färe (1988), "Efficiency Gains from Addition of Technologies: A Nonparametric Approach" in W. Eichhorn, ed., *Measurement in Economics* Heidelberg: Physica-Verlag.
3. S. Grosskopf and S. Yaisawarng (1990), "Economies of Scope in the Provision of Local Public Services," *National Tax Journal* 43, pp.61-74.
4. R. Färe, S. Grosskopf and C.A.K. Lovell (1994), *Production Frontiers* Cambridge: Cambridge University Press.
5. L. Johansen (1968), "Production Functions and the Concepts of Capacity," in: *Recherches Récentes sur la Fonction de Production*, Collection Économie Mathématique et Économétrie 2, Namur.

6. A. Barua, P. L. Brockett, W. W. Cooper, H. Deng, B. R. Parker, T. W. Ruefli and A. Whinston (2004), "DEA Evaluations of Long- and Short-run Efficiencies of "Digital vs. Physical" Product Companies," *Socio-Economic Planning Sciences* 38, pp.233-253.
7. B. Gold (1955), *Foundations of Productivity Analysis*, Pittsburgh University Press.
8. R. Färe, S. Grosskopf and E.C. Kokkelenberg (1989a), "Measuring Plant Capacity Utilization and Technical Change: A Nonparametric Approach," *International Economic Review* 30, pp.655-666 and R. Färe, S. Grosskopf and V. Valdmanis (1989b), "Capacity Competition and Efficiency in Hospitals: A Nonparametric Approach," *Journal of Productivity Analysis* 1, pp.123-128.
9. T. Coelli, E. Grifell-Tatjé and S. Perelman (2002), "Capacity Utilization and Profitability: A Decomposition of Short-Run Profit Efficiency," *International Journal of Production Economics* 79, pp.261-278.
10. K. Tone (2005), "A Price-based Capacity Utilization Analysis Using DEA," The 2005 Fall National Conference of Operations Research Society of Japan.
11. B.K Sahoo and K. Tone (2006), "Decomposing Capacity Utilization in Data Envelopment Analysis: An Application to Banks in India," submitted to *European Journal of Operational Research*.

15 A DEA GAME

15.1 INTRODUCTION

This chapter deals with consensus-making among individuals or organizations under multiple criteria environments.

Suppose that four teammates A, B, C and D successfully finished a job and won a prize pool of \$10,000. Their contributions were multifaceted in that *brain*, *bravery* and *labor* were demanded in its achievement. Table 15.1 exhibits their scores in each category on a scale of one to ten. They carried out the job with coordinated efforts and are willing to share the award in proportion to their contribution. There are several methods for this fair division.

Table 15.1. Score Matrix

	Teammate				Sum
	A	B	C	D	
Brain	9	5	4	2	20
Bravery	6	10	5	3	24
Labor	3	6	7	10	26
Total Score	18	21	16	15	70

For example, a simple allotment rule is given by dividing the reward in proportion to each member's total score. Thus, A will get $\$10,000 \times 18/70 = \$2,571$. Likewise, B, C and D will be rewarded respectively by $\$10,000 \times 21/70 = \$3,000$, $\$10,000 \times 16/70 = \$2,286$ and $\$10,000 \times 15/70 = \$2,143$. This represents the division of the \$10,000 prize when brain, bravery and labor are all equal weighted. Other possibilities are also present as illustrated in Table 15.2.

Table 15.2. Division of Reward based on Fixed Weights

	Teammate				Sum	Weights	
	A	B	C	D		Brain vs. Bravery vs. Labor	
Reward	\$2,571	\$3,000	\$2,286	\$2,143	\$10,000		1 : 1 : 1
Reward	\$3,134	\$3,060	\$2,164	\$1,642	\$10,000		3 : 2 : 1
Reward	\$2,055	\$2,945	\$2,397	\$2,603	\$10,000		1 : 2 : 3

There are many other choices of fixed weights depending on the importance of criteria. For example, if we put emphasis on *brain*, as in row 2, A will get more reward but D may not be satisfied with this weight selection. If we evaluate *labor* as the most important factor, D will be rewarded more with possible discontent from A. Table 15.2 summarizes the division based on these fixed weights rules. As we have pointed out in Section 1.5 of Chapter 1, this fixed weight scheme with their values chosen in an a priori manner can pose a difficult problem on the way we choose the weights.

In order to resolve this problem, we can apply a "variable weight" scheme like the one that is inherent in DEA. Suppose that each member is selfish in that he gives the highest weight to the most preferable criterion for him. Let w_1 , w_2 and w_3 be the weights to the three criteria. Then A will try to maximize his expected reward $9w_1 + 6w_2 + 3w_3$ subject to the total purse constraint $20w_1 + 24w_2 + 26w_3 = 10,000$.¹ The optimal weight for A is $w_1^* = 10,000/20 = 500$, $w_2^* = 0$, $w_3^* = 0$ with the maximum reward $R_A^* = 9w_1^* = \$4,500$. Similarly, each member's optimal weights and reward are exhibited in Table 15.3. As will be demonstrated later, the total reward by this scheme usually exceeds the prize pool. The gap is caused by the selfish weight selection behavior of players.

In this chapter we propose a game-theoretic solution to this problem.

15.2 FORMULATION

We generalize the above problem as follows. There are n players and m criteria with the score matrix $X = (x_{ij}) \in R_+^{n \times m}$. It is assumed that the higher the score for a criterion, the better the player is judged to have performed as regards to this criterion. Each player, k , has the right to choose a set of nonnegative

Table 15.3. Optimal Rewards with Optimal Weights

	A	B	C	D	Sum
w_1^*	500	0	0	0	
w_2^*	0	417	0	0	
w_3^*	0	0	385	385	
Reward	\$4,500	\$4,167	\$2,692	\$3,846	\$15,206 (> \$10,000)

weights $\mathbf{w}^k = (w_1^k, w_2^k, \dots, w_m^k)$ that are most preferable to the player. Using the weight \mathbf{w}^k , we define the relative score of player k to the total score as follows:

$$\frac{\sum_{i=1}^m w_i^k x_{ik}}{\sum_{i=1}^m w_i^k (\sum_{j=1}^n x_{ij})}. \quad (15.1)$$

The denominator represents the total score of all players as measured by player k 's weight selection, while the numerator indicates player k 's self evaluation using the same weight selection. Hence, the expression (15.1) demonstrates player k 's relative importance (share) under the weight (or value) selection \mathbf{w}^k . Player k wishes to maximize this ratio by selecting the most preferable weights, and this results in the following fractional program,

$$\begin{aligned} \max_{\mathbf{w}^k} \quad & \frac{\sum_{i=1}^m w_i^k x_{ik}}{\sum_{i=1}^m w_i^k (\sum_{j=1}^n x_{ij})} \\ \text{subject to } & w_i^k \geq 0. \quad (\forall i) \end{aligned} \quad (15.2)$$

Before continuing, we reformulate the problem as follows, without losing generality. We normalize the data set X so that it is row-wise normalized, i.e.,

$$\sum_{j=1}^n x_{ij} = 1. \quad (\forall i)$$

For this purpose, we divide the row (x_{i1}, \dots, x_{in}) by the row-sum $\sum_{j=1}^n x_{ij}$ for $i = 1, \dots, m$. The program (15.2) is not affected by this operation. Then, using the Charnes-Cooper transformation scheme, the nonlinear fractional program (15.2) can be expressed using a linear program as follows:

$$\begin{aligned} c(k) = \max_{\mathbf{w}^k} \quad & \sum_{i=1}^m w_i^k x_{ik} \\ \text{subject to } & \sum_{i=1}^m w_i^k = 1, \quad w_i^k \geq 0. \quad (\forall i) \end{aligned} \quad (15.3)$$

Now, the problem is to maximize the objective (15.3) on the simplex represented by the constraints. Apparently, the optimal solution is given by assigning 1 to the criterion $i(k)$ such that $x_{i(k)} = \max\{x_{ik} | i = 1, \dots, m\}$, and assigning 0 to the weight of the remaining criteria.² We denote this optimal value by $c(k)$.

The $c(k)$ indicates the highest relative score for player k which is obtained by the optimal weight selecting behavior. The optimal weight will generally differ from one player to another.

Proposition 15.1

$$\sum_{k=1}^n c(k) \geq 1. \quad (15.4)$$

Proof : Let the optimal weight for player k be $w_k^* = (w_{1k}^*, \dots, w_{mk}^*)$, i.e., $w_{i(k)k}^* = 1$ and $w_{ik}^* = 0 (\forall i \neq i(k))$. Then we have

$$\sum_{k=1}^n c(k) = \sum_{k=1}^n \sum_{i=1}^m w_{ik}^* x_{ik} = \sum_{k=1}^n x_{i(k)k} \geq \sum_{k=1}^n x_{1k} = 1.$$

The inequality above follows from $x_{i(k)k} \geq x_{1k}$ and the last equality follows from the row-wise normalization. \square

This proposition asserts that, if each player sticks to his egoistic sense of value and insists on getting the portion of the benefit as designated by $c(k)$, the sum of shares usually exceeds 1 and hence $c(k)$ cannot fulfill the role of division (imputation) of the benefit. If eventually the sum of $c(k)$ turns out to be 1, all players will agree to accept the division $c(k)$, since this is obtained by each player's most preferable weight selection. The sum of the $c(k)$ equals one will occur when all players have the same and common optimal weight selection. More concretely, we have the following proposition.

Proposition 15.2 (Nakabayashi and Tone (2006)³)

The equality $\sum_{k=1}^n c(k) = 1$ holds if and only if the score matrix satisfies the condition

$$x_{1k} = x_{2k} = \dots = x_{mk}, \text{ for } k = 1, \dots, n.$$

That is, each player has the same score with respect to the m criteria.

In the above case, only one criterion is needed for describing the game and the division (imputation) proportional to this score is a fair division. However, such situations occur only in rare instances. In the majority of cases, we have $\sum_{k=1}^n c(k) > 1$. We may call this the “egoist’s dilemma” and many societal problems (conflicts) belong to this class.

One might believe that a benefit allocation proportional to $\{c(k)\}$ is one possible solution. However, this allocation is not rational if we admit coalitions among players.

We explain the above process using the example introduced at the beginning. The normalized score matrix is exhibited in Table 15.4. We also normalize the total prize to 1.

Table 15.4. Normalized Score Matrix

	Teammate				
	A	B	C	D	Sum
Brain	0.4500 ^a	0.2500	0.2000	0.1000	1
Bravery	0.2500	0.4167 ^a	0.2083	0.1250	1
Labor	0.1154	0.2308	0.2692 ^a	0.3846 ^a	1
Total Score	0.8154	0.8975	0.6775	0.6096	3

^a: Underline indicates the maximum entry in each column.

The linear program for A as obtained from (15.3) is given by

$$\begin{aligned} c(A) &= \max \quad 0.45w_1 + 0.25w_2 + 0.1154w_3 \\ \text{subject to} \quad w_1 + w_2 + w_3 &= 1 \\ w_1, w_2, w_3 &\geq 0. \end{aligned}$$

The optimal solution is readily obtained as $w_1^* = 1$, $w_2^* = 0$, $w_3^* = 0$ with $c(A) = 0.45$ which is underlined in the table. Other players' optimal solutions are also recorded in the table by underlined numbers in the respective columns. The total gain due to this selfish principle amounts to $0.45 + 0.4167 + 0.2692 + 0.3846 = 1.5205$ which exceeds the total purse of 1 by 52.05% which is equal to the non-normalized case.

Assumption on the game and fair division

We now assume the following agreements among the players, although each "selfish" player sticks to his most preferable weight selection behavior as expressed by the program (15.3),

- (A1) All players agree not to break off the game.
- (A2) All players are willing to negotiate with each other to attain a reasonable and fair division.

Our main objective now is to find a fair division by cooperative game theory.

15.3 COALITION AND CHARACTERISTIC FUNCTION

Any subset S of the player set $N = (1, \dots, n)$ is called a *coalition*. In our example, coalitions are denoted as $\{A, B\}$, $\{A, C, D\}$ etc. We then define the score for coalition S as the sum of each member's score for each criterion,

$$x_i(S) = \sum_{j \in S} x_{ij}. \quad (i = 1, \dots, m) \tag{15.5}$$

This coalition, i.e., coalition S , aims at obtaining the maximum outcome $c(S)$ by solving the following LP,

$$\begin{aligned} c(S) &= \max_{\mathbf{w}} \sum_{i=1}^m w_i x_i(S) \\ \text{subject to } &\sum_{i=1}^m w_i = 1, \quad w_i \geq 0 \quad (\forall i). \end{aligned} \tag{15.6}$$

The $c(S)$, with $c(\emptyset) = 0$, defines a *characteristic function* of the coalition S . Thus, in this way, we have a cooperative game in coalition form with transferable utility represented by (N, c) .

Proposition 15.3 *The characteristic function c is sub-additive, i.e., for any $S \subset N$ and $T \subset N$ with $S \cap T = \emptyset$, we have*

$$c(S \cup T) \leq c(S) + c(T). \tag{15.7}$$

Proof. By renumbering the indexes, we can assume that

$$S = \{1, \dots, h\}, \quad T = \{h+1, \dots, k\} \text{ and } S \cup T = \{1, \dots, k\}.$$

For these sets, it holds that

$$c(S \cup T) = \max_i \sum_{j=1}^k x_{ij} \leq \max_i \sum_{j=1}^h x_{ij} + \max_i \sum_{j=h+1}^k x_{ij} = c(S) + c(T). \quad \square$$

We also have the following proposition.

Proposition 15.4 $c(N)=1$.

We enumerate all coalitions and their characteristic functions in the above 4 persons case (except for the single player case in Table 15.4) in Tables 15.5 and 15.6 where the characteristic function values are underlined.

Table 15.5. Coalition and Characteristic Function (1)

Coalition	{A,B}	{A,C}	{A,D}	{B,C}	{B,D}	{C,D}
Brain	<u>0.7000^a</u>	<u>0.6500^a</u>	<u>0.5500^a</u>	0.4500	0.3500	0.3000
Bravery	0.6667	0.4583	0.3750	<u>0.6250^a</u>	0.5417	0.3333
Labor	0.3462	0.3846	0.5000	0.5000	<u>0.6154^a</u>	<u>0.6538^a</u>

^a: Underline indicates the characteristic function in each coalition.

Table 15.6. Coalition and Characteristic Function (2)

Coalition	{A,B,C}	{A,B,D}	{A,C,D}	{B,C,D}	{A,B,C,D}
Brain	<u>0.9000^a</u>	<u>0.8000^a</u>	0.7500	0.5500	<u>1^a</u>
Bravery	0.8750	0.7917	0.5833	0.7500	<u>1^a</u>
Labor	0.6154	0.7308	<u>0.7692^a</u>	<u>0.8846^a</u>	<u>1^a</u>

^a: Underline indicates the characteristic function in each coalition.

15.4 SOLUTION

We first evaluate each member's marginal contribution in coalitions and then introduce the "Shapley value" as a solution to this problem.

15.4.1 Coalition and Individual Contribution

In Table 15.5, the coalition {A,B} has the maximum gain 0.7, whereas A and B have respectively the maximum gains 0.45 and 0.4167 as exhibited in Table 15.4. Thus, the marginal contribution of A to the coalition {A,B} can be estimated $c(\{A, B\}) - c(\{B\}) = 0.7 - 0.4167 = 0.2833$ and, for B, it is $c(\{A, B\}) - c(\{A\}) = 0.7 - 0.45 = 0.25$ where the coalition score $c(\{A, B\})$ is obtained from Table 15.5. Turning to Table 15.6, the role of A in the coalition {A,B,C} is $c(\{A, B, C\}) - c(\{B, C\}) = 0.9 - 0.625 = 0.275$.

Table 15.7 exhibits each member's marginal contribution to all coalitions.

15.4.2 The Shapley Value

We have discussed the marginal contribution of player k in coalition S as $c(S) - c(S - \{k\})$. It is reasonable to evaluate k 's contribution to the entire game as the average of marginal contribution to coalitions which include k . With regard to this scheme, the Shapley value⁴ is the representative imputation of cooperative game.

In our DEA game, it goes as follows. We introduce "ordering" in forming a coalition. For example, suppose that A is the first comer to the coalition and B follows him, and further C and D join the coalition in this order. We denote this ordered coalition by $A \leftarrow B \leftarrow C \leftarrow D$. In this coalition D's marginal contribution is evaluated as

$$c(\{A, B, C, D\}) - c(\{A, B, C\}) = 1 - 0.9 = 0.1.$$

C has marginal contribution (independent of D) as taken from Tables 15.6 and 15.5,

$$c(\{A, B, C\}) - c(\{A, B\}) = 0.9 - 0.7 = 0.2.$$

Table 15.7. Each Member's Marginal Contribution to Coalitions

Coalition	A	B	C	D
{A}	0.45	0	0	0
{B}	0	0.4167	0	0
{C}	0	0	0.2692	0
{D}	0	0	0	0.3846
{A,B}	0.2833	0.25	0	0
{A,C}	0.3808	0	0.2	0
{A,D}	0.1654	0	0	0.1
{B,C}	0	0.3558	0.2083	0
{B,D}	0	0.2308	0	0.1987
{C,D}	0	0	0.2692	0.3846
{A,B,C}	0.275	0.25	0.2	0
{A,B,D}	0.1846	0.25	0	0.1
{A,C,D}	0.1154	0	0.2192	0.1192
{B,C,D}	0	0.2308	0.2692	0.2596
{A,B,C,D}	0.1154	0.2308	0.2	0.1

B's marginal contribution, as taken from Tables 15.5 and 15.3, is

$$c(\{A, B\}) - c(\{A\}) = 0.7 - 0.45 = 0.25.$$

Lastly, A's marginal contribution is

$$c(\{A\}) - c(\{\emptyset\}) = 0.45 - 0 = 0.45.$$

Thus, we have a decomposition of marginal contribution in the ordered coalition A←B←C←D,

$$1 = c(\{A, B, C, D\}) = 0.45(A) + 0.25(B) + 0.2(C) + 0.1(D).$$

The first row of Table 15.8 displays this decomposition. Similarly, we count all orderings (permutations) of players and evaluate their marginal contribution to the coalition. See Table 15.8. We assume that all permutations occur with equal probability. Thus, we have what is referred to as the "Shapley value" of each player as the average of the marginal contributions exhibited in the bottom row of Table 15.8. The division of the total purse \$10,000 is recorded in Table 15.9 and differs greatly from the naive shares based on the total score exhibited in Table 15.2.

We now introduce a well-known formula for describing the Shapley value $\phi_k(c)$ of player k by means of the characteristic function $c(S)$ as follows.

$$\phi_k(c) = \frac{1}{n!} \sum_{S: k \in S \subset N} (s-1)!(n-s)! \left\{ c(S) - c(S - \{k\}) \right\}, \quad (15.8)$$

Table 15.8. The Shapley Value

Coalition	A	B	C	D	Sum
A \leftarrow B \leftarrow C \leftarrow D	0.45	0.25	0.2	0.1	1
A \leftarrow B \leftarrow D \leftarrow C	0.45	0.25	0.2	0.1	1
A \leftarrow C \leftarrow B \leftarrow D	0.45	0.25	0.2	0.1	1
A \leftarrow C \leftarrow D \leftarrow B	0.45	0.2308	0.2	0.1192	1
A \leftarrow D \leftarrow B \leftarrow C	0.45	0.25	0.2	0.1	1
A \leftarrow D \leftarrow C \leftarrow B	0.45	0.2308	0.2192	0.1	1
B \leftarrow A \leftarrow C \leftarrow D	0.2833	0.4167	0.2	0.1	1
B \leftarrow A \leftarrow D \leftarrow C	0.2833	0.4167	0.2	0.1	1
B \leftarrow C \leftarrow A \leftarrow D	0.275	0.4167	0.2083	0.1	1
B \leftarrow C \leftarrow D \leftarrow A	0.1154	0.4167	0.2083	0.2596	1
B \leftarrow D \leftarrow A \leftarrow C	0.1846	0.4167	0.2	0.1987	1
B \leftarrow D \leftarrow C \leftarrow A	0.1154	0.4167	0.2692	0.1987	1
C \leftarrow A \leftarrow B \leftarrow D	0.3808	0.25	0.2692	0.1	1
C \leftarrow A \leftarrow D \leftarrow B	0.3808	0.2308	0.2692	0.1192	1
C \leftarrow B \leftarrow A \leftarrow D	0.275	0.3558	0.2692	0.1	1
C \leftarrow B \leftarrow D \leftarrow A	0.1154	0.3558	0.2692	0.2596	1
C \leftarrow D \leftarrow A \leftarrow B	0.1154	0.2308	0.2692	0.3846	1
C \leftarrow D \leftarrow B \leftarrow A	0.1154	0.2308	0.2692	0.3846	1
D \leftarrow A \leftarrow B \leftarrow C	0.1654	0.25	0.2	0.3846	1
D \leftarrow A \leftarrow C \leftarrow B	0.1654	0.2308	0.2192	0.3846	1
D \leftarrow B \leftarrow A \leftarrow C	0.1846	0.2308	0.2	0.3846	1
D \leftarrow B \leftarrow C \leftarrow A	0.1154	0.2308	0.2692	0.3846	1
D \leftarrow C \leftarrow A \leftarrow B	0.1154	0.2308	0.2692	0.3846	1
D \leftarrow C \leftarrow B \leftarrow A	0.1154	0.2308	0.2692	0.3846	1
Shapley Value	0.2584	0.2925	0.2311	0.2180	1

Table 15.9. Division of Reward based on Shapley Value

	Teammate				Sum
	A	B	C	D	
Reward	\$2,584	\$2,925	\$2,311	\$2,180	\$10,000

where s is the number of members in the coalition S . If a subset $S(\subset N)$ includes player k , k 's marginal contribution is obtained as $c(S) - c(S - \{k\})$. We count the occurrence of this contribution in all permutations of the players including k . To see what is involved in arriving at the Shapley value let us assume that the ordered S includes k in the “last comer” position so that it is preceded by $\alpha_1, \dots, \alpha_{s-1}$, and k 's successors are $\beta_1, \dots, \beta_{n-s}$. This pattern

appears $(s - 1)!(n - s)!$ times in the all permutations $n!$. Hence, if we assume that the all permutations (orderings) occur with equal probability, k 's Shapley value can be expressed by the formula for $\phi_k(c)$ in (15.8)

15.5 DEA MIN GAME

We now simplify matters by turning to DEA. For this purpose, we define a DEA *min* game by replacing *max* by *min* in (15.6) as follows,

$$\begin{aligned} d(S) &= \min_w \sum_{i=1}^m w_i x_i(S) \\ \text{subject to } &\sum_{i=1}^m w_i = 1, \quad w_i \geq 0. \quad (\forall i). \end{aligned} \tag{15.9}$$

The optimal value $d(S)$ is the minimum value of the division that the coalition S can expect from the game. Using this optimal solution, we next define a DEA *min* game (N, d) . This game is *super-additive*, i.e.,

$$d(S \cup T) \geq d(S) + d(T), \quad \forall S, T \subset N \text{ with } S \cap T = \emptyset \tag{15.10}$$

Thus, this game starts from $d(k) > 0$ ($k = 1, \dots, n$) and enlarges the gains by coalitions until the grand coalition N emerges with $d(N) = 1$.

Between the games (N, d) and (N, c) we have the following relationship.

Proposition 15.5

$$d(S) + c(N - S) = 1. \quad \forall S \subset N \tag{15.11}$$

Proof: By renumbering the indexes, $N = \{1, \dots, n\}$, we can assume that

$$S = \{1, \dots, h\} \text{ and } N - S = \{h + 1, \dots, n\}.$$

For these two sets, it holds that

$$\begin{aligned} d(S) + c(N - S) &= \min_i \sum_{j=1}^h x_{ij} + \max_i \sum_{j=h+1}^n x_{ij} \\ &= \min_i \left(\sum_{j=1}^n x_{ij} - \sum_{j=h+1}^n x_{ij} \right) + \max_i \sum_{j=h+1}^n x_{ij} \\ &= \min_i \left(1 - \sum_{j=h+1}^n x_{ij} \right) + \max_i \sum_{j=h+1}^n x_{ij} \\ &= 1 - \max_i \sum_{j=h+1}^n x_{ij} + \max_i \sum_{j=h+1}^n x_{ij} = 1. \quad \square \end{aligned}$$

Furthermore, we have the following theorem.

Theorem 15.1 (Nakabayashi and Tone (2006)) *The Shapley values of the DEA max and min games (N, c) and (N, d) are the same.*

Proof. For all $k \in N$ we have

$$\begin{aligned}\phi_k(c) &= \frac{1}{n!} \sum_{S:k \in S \subset N} (s-1)!(n-s)! \left\{ c(S) - c(S - \{k\}) \right\} \\ &= \frac{1}{n!} \sum_{S:k \in S \subset N} (s-1)!(n-s)! \left[\left\{ 1 - d(N-S) \right\} - \left\{ 1 - d(N-S+\{k\}) \right\} \right] \\ &= \frac{1}{n!} \sum_{S:k \in S \subset N} (s-1)!(n-s)! \left\{ d(N-S+\{k\}) - d(N-S) \right\}.\end{aligned}$$

By replacing $S' = N - S + \{k\}$ and defining s' as the number of members in the coalition S' , the last term becomes:

$$\begin{aligned}&= \frac{1}{n!} \sum_{S':k \in S' \subset N} (s'-1)!(n-s')! \left\{ d(S') - d(S' - \{k\}) \right\} \\ &= \phi_k(d),\end{aligned}$$

where $\phi_k(d)$ is the Shapley value defined in (15.8). \square

This theorem demonstrates an interesting feature of the Shapley value in DEA games. Although the *max* game represents selfish (bullish) behavior of players and the *min* game corresponds to self-sacrificing (bearish) behavior, the Shapley solutions are the same in this cooperative game, i.e., $\phi_k(d) = \phi_k(c)$.

15.6 SUMMARY OF CHAPTER 15

In this chapter we have introduced a consensus-making method in a multiple criteria environment using a combination of DEA and cooperative game theory. It is demonstrated that both DEA *max* and *min* games have the same Shapley value.

Problems like the one exemplified in this chapter are usually solved by means of (among others)

- conventional custom
- a single criterion decision or
- a fixed weights rule.

These approaches are not always “rational.”

The proposed scheme has diverse applications in areas, such as,

- cost or burden sharing in international organizations, e.g., United Nations, NATO, UNESCO and so forth,

- research grant allocation to applicants by a foundation, and
- resource distribution for R&D.

These problems are multifaceted and should be solved in a cooperative framework. The DEA game proposed in this chapter can be a promising method for solving these important problems and will open a new dimension to cooperative game theory.

15.7 NOTES AND SELECTED BIBLIOGRAPHY

The contents of this chapter are based on Nakabayashi and Tone (2006) where the DEA game is further developed and discussed in detail, e.g., the core, the nucleolus, the benefit-cost game and the relationship with Owen's linear production games.⁵ See also Nakabayashi (2005).⁶

15.8 PROBLEM SUPPLEMENT FOR CHAPTER 15

Problem 15.1

Tables 15.10, 15.11 and 15.12 exhibit the characteristic function values for the *min* game case of the 4-person game as presented in the text. They correspond to Tables 15.4, 15.5 and 15.6 of the *max* game, respectively.

Table 15.10. Single Players' Values for the Min Game Case

	A	B	C	D	Sum
Brain	0.4500	0.2500	<u>0.2000^a</u>	<u>0.1000^a</u>	1
Brave	0.2500	0.4167	0.2083	0.1250	1
Labor	<u>0.1154^a</u>	<u>0.2308^a</u>	0.2692	0.3846	1

^a: Underline indicates the characteristic function for the *min* game in each player.

Table 15.11. Coalitions' Values for the Min Game Case (1)

Coalition	{A,B}	{A,C}	{A,D}	{B,C}	{B,D}	{C,D}
Brain	0.7000	0.6500	0.5500	<u>0.4500^a</u>	<u>0.3500^a</u>	<u>0.3000^a</u>
Brave	0.6667	0.4583	<u>0.3750^a</u>	0.6250	0.5417	0.3333
Labor	<u>0.3462^a</u>	<u>0.3846^a</u>	0.5000	0.5000	0.6154	0.6538

^a: Underline indicates the characteristic function for the *min* game in each coalition.

Table 15.12. Coalitions' Values for the Min Game Case (2)

Coalition	{A,B,C}	{A,B,D}	{A,C,D}	{B,C,D}	{A,B,C,D}
Brain	0.9000	0.8000	0.7500	<u>0.5500^a</u>	<u>1^a</u>
Brave	0.8750	0.7917	<u>0.5833^a</u>	0.7500	<u>1^a</u>
Labor	<u>0.6154^a</u>	<u>0.7308^a</u>	0.7692	0.8846	<u>1^a</u>

^a: Underline indicates the characteristic function for the *min* game in each coalition.

Assignment: Using the above characteristic function, calculate the Shapley value and confirm Theorem 15.1 that the Shapley value of the *min* game coincides with that of the *max* game in Table 15.8.

Suggested Response: Similar to Table 15.8, we have the Shapley value at the bottom row of Table 15.13. The Shapley values in Tables 15.8 and 15.13 are (0.2584, 0.2925, 0.2311, 0.218) and they are the same.

Problem 15.2

Introduction: A shopping mall association made the following agreement on the arcade maintenance fee: “Every shop facing the arcade street has to pay a monthly fee. The method for arriving at this fee for each shop was discussed and approved at the general meeting. Share of cost was determined based on parameters such as category of business, the size of the shop, the number of employees, and so on.”

Suppose that only three shops A, B and C face the arcade street and they have to determine share of the arcade maintenance fee based on three criteria such as the profits, the number of employees and the sizes of the three shops, as exhibited in Table 15.14.

Assignment: Part I: How is share of cost determined? Using the DEA game, propose the Shapley solution. Part II: Suppose that all the shops agree that Criterion 1 (profits) is more important than Criteria 2 and 3 (employee and size). In this situation, we can apply the *assurance region (AR) method* as explained in Section 6.2 of Chapter 6. Add appropriate constraints to original LP formulation (15.6) or (15.9). Part III: Calculate the Shapley value in the above AR case and compare it with the solution proposed in Part I.

Suggested Responses: Part I: Tables 15.15 and 15.16 exhibit the characteristic function values for the *min* game case.

Using the above characteristic function, we have the Shapley value in the last row of Table 15.17 that is (0.328, 0.262, 0.410).

Table 15.13. The Shapley Value for the Min Game Case

Coalition	A	B	C	D	Sum
A \leftarrow B \leftarrow C \leftarrow D	0.1154	0.2308	0.2692	0.3846	1
A \leftarrow B \leftarrow D \leftarrow C	0.1154	0.2308	0.2692	0.3846	1
A \leftarrow C \leftarrow B \leftarrow D	0.1154	0.2308	0.2692	0.3846	1
A \leftarrow C \leftarrow D \leftarrow B	0.1154	0.4167	0.2692	0.1987	1
A \leftarrow D \leftarrow B \leftarrow C	0.1154	0.3558	0.2692	0.2596	1
A \leftarrow D \leftarrow C \leftarrow B	0.1154	0.4167	0.2083	0.2596	1
B \leftarrow A \leftarrow C \leftarrow D	0.1154	0.2308	0.2692	0.3846	1
B \leftarrow A \leftarrow D \leftarrow C	0.1154	0.2308	0.2692	0.3846	1
B \leftarrow C \leftarrow A \leftarrow D	0.1654	0.2308	0.2192	0.3846	1
B \leftarrow C \leftarrow D \leftarrow A	0.45	0.2308	0.2192	0.1	1
B \leftarrow D \leftarrow A \leftarrow C	0.3808	0.2308	0.2692	0.1192	1
B \leftarrow D \leftarrow C \leftarrow A	0.45	0.2308	0.2	0.1192	1
C \leftarrow A \leftarrow B \leftarrow D	0.1846	0.2308	0.2	0.3846	1
C \leftarrow A \leftarrow D \leftarrow B	0.1846	0.4167	0.2	0.1987	1
C \leftarrow B \leftarrow A \leftarrow D	0.1654	0.25	0.2	0.3846	1
C \leftarrow B \leftarrow D \leftarrow A	0.45	0.25	0.2	0.1	1
C \leftarrow D \leftarrow A \leftarrow B	0.2833	0.4167	0.2	0.1	1
C \leftarrow D \leftarrow B \leftarrow A	0.45	0.25	0.2	0.1	1
D \leftarrow A \leftarrow B \leftarrow C	0.275	0.3558	0.2692	0.1	1
D \leftarrow A \leftarrow C \leftarrow B	0.275	0.4167	0.2083	0.1	1
D \leftarrow B \leftarrow A \leftarrow C	0.3808	0.25	0.2692	0.1	1
D \leftarrow B \leftarrow C \leftarrow A	0.45	0.25	0.2	0.1	1
D \leftarrow C \leftarrow A \leftarrow B	0.2833	0.4167	0.2	0.1	1
D \leftarrow C \leftarrow B \leftarrow A	0.45	0.25	0.2	0.1	1
Shapley Value	0.2584	0.2925	0.2311	0.2180	1

Table 15.14. Data of 3 Shops

Shop	A	B	C	Sum
Criterion 1: Profits	10	5	7	22
Criterion 2: Employee	9	16	20	45
Criterion 3: Size	20	10	30	60

Part II: Using the assurance region method, the following constraints are to be added to the original LP formulation.

$$U_i \geq \frac{w_i}{w_1} \geq L_i \quad (i = 2, 3),$$

Table 15.15. Single Players' Values of the Market Arcade Game

	A	B	C	Sum
Profits	0.4545	0.2273	<u>0.3182^a</u>	1
Employee	<u>0.2000^a</u>	0.3556	0.4444	1
Size	0.3333	<u>0.1667^a</u>	0.5000	1

^a: Underline indicates the characteristic function for the *min* game in each player.

Table 15.16. Coalitions' Values of the Market Arcade Game

Coalition	{A,B}	{A,C}	{B,C}	{A,B,C}
Profits	0.6818	0.7727	<u>0.5455^a</u>	<u>1^a</u>
Employee	0.5556	<u>0.6444^a</u>	0.8000	<u>1^a</u>
Size	<u>0.5000^a</u>	0.8333	0.6667	<u>1^a</u>

^a: Underline indicates the characteristic function for the *min* game in each coalition.

Table 15.17. The Shapley Value of the Market Arcade Game

Coalition	A	B	C	Sum
A←B←C	0.2	0.3	0.5	1
A←C←B	0.2	0.3556	0.4444	1
B←A←C	0.3333	0.1667	0.5	1
B←C←A	0.4545	0.1667	0.3788	1
C←A←B	0.3262	0.3556	0.3182	1
C←B←A	0.4545	0.2273	0.3182	1
Shapley Value	0.328	0.262	0.410	1

In this problem, we require $U_i = 1$ ($i = 2, 3$) and do not require the lower bounds. That is, the constraints $w_1 \geq w_2$ and $w_1 \geq w_3$ are added and the original linear program is modified as follows:

$$d(S) = \min_{\mathbf{w}} \sum_{i=1}^3 w_i x_i(S) \quad (15.12)$$

subject to

$$\begin{aligned} \sum_{i=1}^3 w_i &= 1 \\ w_1 &\geq w_2, \quad w_1 \geq w_3 \end{aligned}$$

$$w_1, w_2, w_3 \geq 0.$$

Part III: For convenience, we here solve the LP problem (15.12) as follows:

$$d(S) = \min \left\{ x_1(S), \frac{x_1(S) + x_2(S)}{2}, \frac{x_1(S) + x_3(S)}{2}, \frac{x_1(S) + x_2(S) + x_3(S)}{3} \right\}. \quad (15.13)$$

Table 15.18 exhibits the characteristic function values for the AR DEA game case.

Table 15.18. Characteristic Function Values subject to the AR Constraints

Player/Coalition	$x_1(S)$	$\frac{x_1(S)+x_2(S)}{2}$	$\frac{x_1(S)+x_3(S)}{2}$	$\frac{x_1(S)+x_2(S)+x_3(S)}{3}$
A	0.4545	<u>0.3273^a</u>	0.3939	0.3293
B	0.2273	0.2914	<u>0.197^a</u>	0.2498
C	<u>0.3182^a</u>	0.3813	0.4091	0.4209
{A,B}	0.6818	0.6187	0.5909	<u>0.5791^a</u>
{A,C}	0.7727	<u>0.7086^a</u>	0.803	0.7502
{B,C}	<u>0.5455^a</u>	0.6727	0.6061	0.6707
{A,B,C}	1 ^a	1 ^a	1 ^a	1 ^a

^a: Underline indicates the characteristic function for the AR case in each player/coalition.

Using the above characteristic function, we have the Shapley value at the last row of Table 15.19 that is (0.389, 0.243, 0.368). Shop C is ranked higher than Shop A in the solution proposed by Part I (before the AR method is applied), whereas Shop A ranks first in this AR case because Criterion 1 (profits) is considered to be more important (Shop A makes the highest profit).

Table 15.19. The Shapley Value for the AR Case

Coalition	A	B	C	Sum
A \leftarrow B \leftarrow C	0.3273	0.2518	0.4209	1
A \leftarrow C \leftarrow B	0.3273	0.2814	0.3813	1
B \leftarrow A \leftarrow C	0.3821	0.197	0.4209	1
B \leftarrow C \leftarrow A	0.4545	0.197	0.3485	1
C \leftarrow A \leftarrow B	0.3904	0.2914	0.3182	1
C \leftarrow B \leftarrow A	0.4545	0.2273	0.3182	1
Shapley Value	0.389	0.243	0.368	1

Notes

1. These same weights are to be assigned to the similar contributions of brain, bravery and labor to other participants.
2. We assume that there are no alternate optima. If such alternatives present we can choose one arbitrarily and assign zero weights to the remaining criteria.
3. K. Nakabayashi and K. Tone (2006), "Egoist's Dilemma: A DEA Game," *Omega* 34, pp.135-148.
4. L.S. Shapley (1953), "A Value for n -Person Games," in H.W. Kuhn and A.W. Tucker, eds., *Contributions to the Theory of Games* Vol. II (Annals of Mathematics Studies, 28) pp.307-317, Princeton University Press.
5. G. Owen (1975), "On the Core of Linear Production Games," *Mathematical Programming* 9, pp.358-370.
6. K. Nakabayashi (2005), "Egoist's Dilemma: A DEA Game," Ph.D. Thesis, National Graduate Institute for Policy Studies.

16 MULTI-STAGE USE OF PARAMETRIC AND NON-PARAMETRIC MODELS

16.1 INTRODUCTION

In this chapter we first cover statistical regression vs. DEA comparisons and then move on to statistical regression combinations. We start with some of the earliest efforts after which we turn to some of the more recent developments in what is still an expanding area of research on the use of these two approaches to evaluating performances. The development starts with OLS (Ordinary Least Squares) and then moves on to the so called “thick frontier” models followed by the formulations of Aigner & Chu (AC), COLS (Corrected OLS), and SFA (Stochastic Frontier Analysis) regressions. We then turn attention to efforts that combine DEA and regression approaches in attempts to extend the capacities of both. We start with a two-stage DEA-OLS combination and then extend this to multi-stage combinations of DEA and SFA in the latter part of this chapter. Finally, we introduce a case study dealing with a three-stage approach to measurements of Japanese banking efficiencies over the years 1997-2001.

16.2 OLS REGRESSIONS

The statistical regression models we use take the following form:

$$y = f(\mathbf{x}) + \varepsilon, \quad (16.1)$$

where y , the regressand, is a scalar, and \mathbf{x} is a vector of regressor variables. This is referred to as a “fixed effects” model with the random errors, ε , assigned entirely to y , the regressand. Thus, it is to be distinguished from so-called “stochastic models” in which some or all of the regressors also have random components.

The regressions we deal with all have known (or assumed) parametric forms and thus differ from DEA which is “non parametric.” Most of the studies we deal with assume that (16.1) takes the form of a Cobb-Douglas function,

$$y = \beta_0 \prod_{i=1}^m x_i^{\beta_i} e^\varepsilon. \quad (16.2)$$

Taking logarithm this is transformed into the following “Cobb-Douglas,” or log-linear, form

$$\ln y = \ln \beta_0 + \sum_{i=1}^m \beta_i \ln x_i + \varepsilon. \quad (16.3)$$

The parameter values are then estimated by OLS which yields BLUE (Best Linear Unbiased Estimates) on the assumption of iid (independent identically distributed) observations with constant mean μ and variance, σ^2 .

Early studies usually took the form of comparisons between such approaches and DEA with mixed results. See Berger and Humphrey (1997)¹ for an extensive survey of such studies. A problem arises in such studies because the orientations differ. OLS regressions are oriented toward *central tendency* estimates that yield expected values. They thus provide estimates of the conditional *averages* that are expected to prevail because of past performances from the collection of observed values. DEA, on the other hand, is directed to optimal estimates for *each* individual entity (=DMU) represented in a collection of observations, with their performances being evaluated relative to *frontiers* formed from the performances that the data show to be possible of attainment. Thus, by contrast, DEA is individually rather than average oriented and deals with frontiers rather than central tendencies.

Various versions of these regression based approaches will next be described along with simulations and similar kinds of studies that make comparisons possible with *known* parametric values and efficient performances.

16.3 MODIFICATION AND EXTENSIONS

An early attempt to modify these OLS approach is called “thick frontiers.” This approach uses two fixed effects regressions which are estimated from data for the highest and lowest quartile. Deviations from each regression are assumed to represent random errors while differences in the estimates from these two regressions represent inefficiencies.

Deficiencies in this thick frontier approach include a use of only 50% of the data. They also include the possibility of crossovers, which can occur when a coefficient of the upper regression is negative and a corresponding coefficient of

the lower regression is positive. See Berger and Humphrey (1997)² for further comments and examples of uses of such “thick frontier” regressions.

Another early approach, due to Aigner and Chu (1986),³ takes the form of a linear programming model as follows,

$$\begin{aligned} & \min \sum_{j=1}^n \varepsilon_j \\ & \text{subject to} \\ & \ln \beta_0 + \sum_{i=1}^m \beta_i \ln x_{ij} - \ln y_j = \varepsilon_j, \quad j = 1, \dots, n \\ & \varepsilon_j, \beta_i \geq 0 \quad \forall i, j, \end{aligned} \tag{16.4}$$

where $j = 1, \dots, n$ refers to the observations.

Schmidt (1976)⁴ showed that the A&C linear programming model is equivalent to a maximum likelihood estimations model if the errors, ε_j , are assumed to have a one-sided exponential distribution. In the operations research literature it would be identified as a “goal programming model” with only one-sided deviations allowed. See Cooper (2005)⁵ for a discussion of relations between goal programming and DEA. Since goal programming had its origins in the use of MAD (minimum absolute deviation) regressions, this would be the way it would be identified in the statistics literature. See Charnes, *et al.* (1957).⁶ This means that the normal distribution assumption may be dropped. It also ameliorates the problem of outlier sensitivity that is present in uses of least squares and provides a more robust measure that is superior to the OLS estimate for a wider class of statistical distributions. See Basset and Koenker (1978).⁷ See also Dielman and Pfaffenbuger (1982).⁸

Unfortunately the A&C model has been found to perform poorly under conditions, such as heteroscedasticity, that are likely to be encountered in managerial applications. See Banker *et al.* (2004).⁹ We therefore turn to COLS (Corrected OLS) and one of the many simulation studies that have now been conducted to compare DEA with alternative parametric estimate under widely varying conditions.

As formulated by Richmond (1974)¹⁰ the estimation process begins with standard OLS formulations to take advantage of its attractive (BLUE) estimating properties. In a succeeding step the intercept term is adjusted upward by the amount of the maximum residual value for the thus estimated regression in order to derive an estimated production frontier value (Greene, 1980).¹¹

The above is referred to as the COLS-R (=Residual) method. Another approach takes the form of the COLS-M (=Moments) method which shifts the intercept term upward by the estimated bias of the error term, ε . This is accomplished by use of the second and third moments of the OLS residuals (Olson *et al.*, 1980).¹²

Extensively used in simulation studies, partly because of its simplicity, COLS has generally been found to perform less well than the non-parametric DEA. However, the study in Banker *et al.* (2004)¹³ found that the best performer, by

far, was a combination in which DEA scores were incorporated in a log linear regression. This combination approach outperformed not only DEA, but also A&C and COLS, as well as the “composed error” method that we describe in the next section.

16.4 STOCHASTIC FRONTIER ANALYSIS AND COMPOSED ERROR MODELS

The composed error regression approach is also referred to as SFA (Stochastic Frontier Analysis). See Kumbhakar and Lovell (2000).¹⁴ The term, composed error, arises from decomposing the error term ε into the two components represented by

$$\varepsilon = \nu - \tau \quad (16.5)$$

with $\tau \geq 0$ and ν unrestricted in sign. The motivating principle is to retain the error term to represent “white noise,” which is generally assumed to be normally distributed (as in OLS) and is unrestricted sign, as in the standardized $N(0, \sigma_{\nu u}^2)$. This noise component represents random variations in output, y , due either to (i) errors of observation or (ii) factors outside the control of management. Then τ is used to represent “technical inefficiencies” and assumed to have a distribution such as the exponential or half normal to insure that it produces only non-negative values.

Since the SFA approach treats inputs as given and uses an output-oriented measure of efficiency, the production function to be estimated (in log-form) which replaces (16.3) is

$$\ln y = \ln \beta_0 + \sum_{i=1}^m \beta_i \ln x_i + \nu - \tau. \quad (16.6)$$

Hence

$$\ln y + \tau = \ln \beta_0 + \sum_{i=1}^m \beta_i \ln x_i + \nu. \quad (16.7)$$

Since $\tau \geq 0$ we always have

$$\ln y + \tau \geq \ln y. \quad (16.8)$$

Thus $\tau \geq 0$ represents the output increase that could have been obtained if this amount of inefficiency was eliminated. In other words, $\tau \geq 0$ represents the amount of technical inefficiency. Notice now that, as in (16.7), SFA allows deviation above and below $\ln y + \tau$ to reflect error estimates even in the thus adjusted regression values.

We now employ assumptions that are standard in the SFA literature and assume that τ and ν are statistically independent. Following Jondrow, Lovell, Materov and Schmidt (1982)¹⁵ we estimate τ via

$$\hat{\tau} = - \left[\mu_\tau - \sigma \frac{f^*(\mu/\sigma)}{F^*(-\mu_\tau/\sigma)} \right]$$

where (16.9)

$$\mu_\tau = \frac{\sigma_\tau^2 \varepsilon}{\sigma_\nu^2 + \sigma_\tau^2} \text{ and } \sigma^2 = \frac{\sigma_\nu^2 \sigma_\tau^2}{\sigma_\nu^2 + \sigma_\tau^2},$$

and $f^*(\cdot)$ and $F^*(\cdot)$ respectively represent the standard normal density and cumulative normal distributions.

The efficiency corresponding to the x_i choices can then be estimated from

$$0 \leq e^{-\tau} \leq 1 \quad (16.10)$$

which is equal to unity when $\tau = 0$ and 0 as $\tau \rightarrow \infty$. See the representation of ε in (16.2). Thus, the inefficiencies as well as the statistical errors are all in the regressand, y , and not in any of the x_i .

16.5 DEA AND REGRESSION COMBINATIONS

We now turn to more recent research which moves from DEA *vs.* regression comparisons to DEA *and* regression combinations. One such approach, as described in Arnold *et al.* (1996)¹⁶ takes a two-stage form as follows: Stage one uses DEA to identify efficient vs. inefficient DMUs. Stage two incorporates these results in the form of dummy variables in the corresponding regression. Symbolizing this dummy variable as D , this is accomplished as follows,

$$D = \begin{cases} 1 & : \text{if a DMU is identified as 100\% DEA efficient in Stage One} \\ 0 & : \text{if a DMU is not identified as 100\% DEA efficient in Stage One.} \end{cases} \quad (16.11)$$

The regression we use extends (16.3) in the following manner,

$$\ln y = \ln \beta_0 + \sum_{i=1}^m \beta_i \ln x_i + \delta D + \sum_{i=1}^m \delta_i D \ln x_i + \varepsilon. \quad (16.12)$$

where δ_i and δ represent the statistical estimates associated with these dummy variables and the variable with which it is associated. To see how this affects the coefficient estimates we collect terms in (16.12) to obtain

$$\ln y = \ln \beta_0 + \delta + \sum_{i=1}^m (\beta_i + \delta_i D) \ln x_i + \varepsilon. \quad (16.13)$$

Thus if $D = 0$ there is no effect on the coefficient $\beta_i = (\beta_i + \delta_i D)$ for $\ln x_i$ or on the intercept $\ln \beta_0 + \delta D$. If $D = 1$, however, the coefficient values are augmented or diminished according to whether the corresponding δ_i is positive or negative.

A simulation study undertaken by Bardhan *et al.* (1998, p.273)¹⁷ reported the following results with the use of (16.11):

"In every case the results were the same: (1) Our two-stage DEA-dummy variable approach yielded estimates that did not differ significantly from true parameter values and (2) the customary single-stage [OLS] regression failed to do

this when applied to observations that reflect mixtures of efficient and inefficient behavior. Further, the customary single-stage approaches failed to perform satisfactorily and yielded biased parameter estimates in every case."

"Additional advantages in our Bardhan *et al.* DEA-dummy variable approaches were as follows. In every case the dummy variable adjusted coefficient gave results that conformed to what is to be expected [from economic theory] when technical efficiency is present and, equally important, a failure to conform to these conditions (as determined by DEA) was always associated with behavior that does not conform to what is to be expected when production is technically efficient . . ."

The above simulations were extended by Bardhan *et al.* (1998)¹⁸ to SFA and to the CES (Constant Elasticity of Substitution) production function with similar results. Changes in the mix of inefficient vs. efficient DMUs used in the simulation design did not alter these results. In part this was due to the fact that input as well as output inefficiencies were present in the data, contrary to the assumption in SFA that inefficiencies appear only in the outputs.¹⁹

This simulation experiment could also have been expanded by using DEA to adjust for these input mixes. In any case these DEA-regression (including SFA) combinations have opened new possibilities for research and use which we reflect in detail in the following sections where a SFA-DEA combination is used to provide a great deal of information that is otherwise not possible to come by. See Section 9.3 in Chapter 9 for further details on this study and its results.

16.6 MULTI-STAGE DEA-REGRESSION COMBINATIONS AND ITS APPLICATION TO JAPANESE BANKING

We now turn to work that builds on a three-stage analysis introduced by Fried *et al.* (2002)²⁰ which distinguishes environmental effects such as the state of the economy, from both other types of (managerial only) inefficiencies and random errors. For this purpose we use a version of the Slacks Based Measure (SBM) discussed earlier in Section 4.4 in order to include all of the inefficiencies—in contrast to the radial measure model used by Fried *et al.* For this purpose we use the following version of an output (only) oriented version of SBM since the environmental variables we deal with are all only outputs.

16.6.1 Introduction

Thus far, we have maintained an important assumption that the data to be used in DEA models is assumed to be deterministic without any measurement errors. In modeling real world problems, this has often been regarded as a strong assumption. This section examines the impact of such assumptions on measured DEA efficiency as well as explores possible ways to relax this assumption.

Except for the windows analysis and the Malmquist Index models, the DEA models covered thus far are intended for cross sectional analysis, i.e., evaluating

the performance of the DMU under consideration within a single period of time. The influence on measured efficiency of the deterministic assumption is then less obvious.

However, in conducting a comparative statics analysis, the problem surfaces, and, in some cases, it may reveal a serious bias to our interpretation of the empirical results. For instance, in Figure 16.1, we plot the mean efficiency score of Japanese banking over the period from fiscal year (FY) 1997 through 2001, which exhibits volatile patterns over time. It appears to be at odds with the general perception of learning-by-doing due to Arrow,²¹ because it implies that, on average, the banks in question conducted the same business activities in a volatile and decreasingly competent way over time. Putting this matter differently, the bankers did not learn from their experience.

Can we correctly draw a conclusion from this result that Japanese banks experienced a decline in productivity over the sample period? No, for the reason that many important factors were not controlled for in such preliminary analysis. For instance, what account is taken of the behavior of the economy and the market conditions that affect the activities of both the banks and their customers? Examining these results in a broader macroeconomic context, one would take note that there were systematic downward trends in the macroeconomic and market conditions in Japan over this period (see, for example, IMF²²). Comparing Figure 16.1 with the figures portraying the macroeconomic and market conditions in the IMF report, it is obvious that the banks enjoyed better performance in good years, while suffering from productivity setbacks in bad years. The coincidence in the cyclical pattern as well as the magnitudes in the peaks and the troughs naturally gives rise to a conjecture, i.e., the volatile and worsening performance may not indicate a decline in banking productivity; instead, it could reflect that the influences of unfavorable operating environment and random shocks to this sector (for instance, stochastic bank runs) were mixed into this performance evaluation. Thus, accepting such empirical evidence at its face value could sometimes be seriously misleading.

DEA researchers have long been aware of environmental influences and other non-discretionary (sometimes causal) factors on DEA efficiency. Works dealing with such problem range from the single-stage non-discretionary and categorical models as in Banker and Morey²³ (also see Chapter 7), to the two-stage approach advanced in Timmer.²⁴ However, measurement errors were left unattended in these early efforts.

Recently, Fried *et al.*²⁵ argue that managerial incompetence alone does not suffice to explain the variation in efficiency, since environmental effects and statistical noise could also exert some influence on a producer's performance. For instance, a seemingly inefficient producer could have used more inputs or failed to produce the optimal level of outputs simply because of some exogenous and unexpected factors at the time of production planning. Properly controlling for such impact, this producer could be efficient, as in fact was the case. Thereby, they suggested the slacks be examined further with econometric techniques to estimate the impact of the environmental effects and statistical noise. Such im-

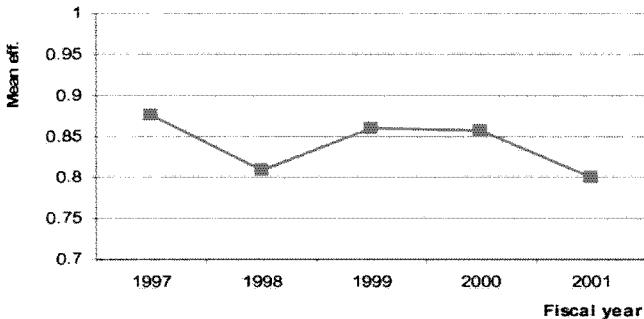


Figure 16.1. Downward Trend in Mean Measured Efficiency by Static DEA Model

pact should be adjusted for when present so as to obtain more valid efficiency estimates for a producer.

To that end, Fried *et al.* used three stages of analyses. In the first stage, they used DEA to estimate slacks.²⁶ In the second stage, they regressed the slacks onto a set of environmental dummies using the stochastic frontier analysis (SFA) to decompose the slacks into three parts, namely, the effects of environmental factors, statistical noise and managerial inefficiency. Using the thus obtained estimates, they proceeded to adjust input data by moving all firms to an artificial common operation environment with maximum impacts of data noise and environmental factors. In the final stage, they again used DEA to estimate efficiency scores. As such, they found significantly different measured efficiencies, which they claimed provided better estimates of the true managerial efficiency.

The idea points in an important direction of improvement. Some recent researches have attempted to improve on this pioneering work. (See, for example, Liu and Tone²⁷ and Avkiran and Rowlands²⁸). Hereinafter, we will introduce a slacks-based multistage framework to simultaneously control for the impacts of environmental effects and statistical noise in each cross section. The adjusted data can then be used for performance evaluation in the context of either a comparative statics study or a Malmquist index analysis.

16.6.2 The Multistage Framework

In this section, we will show how slacks estimated by the SBM model can be differentially assigned to the impacts of environmental factors, measurement errors and managerial inefficiency.

The SBM Model (1st Stage DEA) We assume that there are n decision-making units (DMU_j , $j = 1, \dots, n$), each using m inputs ($i = 1, \dots, m$) to produce s outputs ($r = 1, \dots, s$). For each DMU_o ($o = 1, \dots, n$) under evaluation, the input (output) vectors are denoted as $\mathbf{x}_o \in R_+^m$ and $\mathbf{y}_o \in R_+^s$. The input and output matrices are defined as $X = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in R_+^{m \times n}$ and $Y = (\mathbf{y}_1, \dots, \mathbf{y}_n) \in R_+^{s \times n}$, respectively. For reasons to be explained in Section 16.6.3, the SBM output-oriented model with the assumption of variable returns to scale (hereinafter as SBM) is formulated as:

$$[\text{SBM}] \quad \min_{\lambda, s^+} \rho_o^* = \frac{1}{1 + \frac{1}{s} \sum_{r=1}^s s_r^+ / y_{ro}} \quad (16.14)$$

$$\text{subject to} \quad x_{io} \geq \sum_{j=1}^n x_{ij} \lambda_j \quad (i = 1, \dots, m) \quad (16.15)$$

$$y_{ro} = \sum_{j=1}^n y_{rj} \lambda_j - s_r^+ \quad (r = 1, \dots, s) \quad (16.16)$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (16.17)$$

$$\lambda \geq \mathbf{0}, \quad s^+ \geq \mathbf{0},$$

where λ is an intensity vector of dimension n , and s^+ is the s -dimensional vectors of output slacks.

We can employ the weighted SBM model (WSBM) in Section 4.4.7 in Chapter 4 if imposition of weights to output items is needed.

Doubly Heteroscedastic SFA Model (2nd Stage SFA) DEA empirical studies typically involve accounting data, which is notoriously inaccurate and sensitive to the change in accounting standards as well as in various price levels that are closely related to macroeconomic and market conditions. Besides, the output shortfalls can be attributable to whether the environment is facilitating or not, as well as to random shocks, such as catastrophic earthquakes and hurricanes and so forth, that may catch managers off guard. Therefore, it is advisable to control for the effects of environmental factors, statistical noise and/or random shocks (hereinafter, the latter two are collectively referred to as “random shocks”). Assume that the observed inefficiency effect s_r^{+*} / y_{ro} in output r is a function of the true, but unobservable, management incompetence, $\hat{s}_r^{+*} / \hat{y}_{ro}$, and the two aforementioned effects as follows:

$$s_r^{+*} / y_{ro} = (\hat{s}_r^{+*} / \hat{y}_{ro}) + f(en) + v, \quad (r = 1, \dots, s) \quad (16.18)$$

where $f(en)$ is the function of environmental factors and v is the effects of random shocks, respectively.

To estimate equation (16.18), we introduce the doubly-heteroscedastic SFA models. Since slacks can be considered as opportunity cost to firms, we employ

the cost frontier model. We follow Battese and Coelli²⁹ to formulate the environmental effects as a Cobb-Douglas function, for computational simplicity, but without loss of generality, as follows:

$$\frac{s_{rjt}^{+*}}{y_{rjt}} = \beta_{ro} + \sum_{k=1}^K \beta_{rk} \ln Z_{kjt} + v_{rjt} + u_{rjt}, \quad (r = 1, \dots, s). \quad (16.19)$$

Here j ($j = 1, \dots, N$) is the cross-sectional identifier and t ($t = 1, \dots, T$) is the time identifier. The dependent variable s_{rjt}^{+*}/y_{rjt} is the normalized optimal slack in output r of DMU $_j$ at time t obtained in the 1st stage.

On the right-hand side, Z contains K observable environmental factors ($k = 1, \dots, K$). β_{ro} is the intercept of the r^{th} equation. β_{rk} describes the average responses of the firms in the sample to the changes in the exogenous variables that enter into equation r . u_{rjt} represents $\hat{s}_r^{+*}/\hat{y}_{ro}$, and $v_{rjt} \sim N[0, \sigma_{v_{rjt}}^2]$ is the classical error term that catches idiosyncratic random shocks. The other standard SFA assumptions are applied here.

To correct for the heteroscedasticity, we follow Greene³⁰ to specify the variance equations of the composed disturbance term as:

$$\sigma_{v_{rjt}}^2 = \sigma_v^2 \times \exp(\delta_r^t W_{jt}) \quad (r = 1, \dots, s) \quad (16.20)$$

$$\sigma_{u_{rjt}}^2 = \sigma_u^2 \times \exp(\gamma_r^t P_{jt}) \quad (r = 1, \dots, s) \quad (16.21)$$

where the matrices W and P contain regressors explaining the variance in the v term and the u term, respectively. When estimating the heteroscedastic model in the u term, only the half normal assumption is suitable. Furthermore, the u term is assumed to have zero underlying mean, since a model with nonzero mean and variation in the variance appears to be not estimable (See Greene³¹). Therefore, we have $u_{rjt} \sim N^+[0, \sigma_{u_{rjt}}^2]$.

As a caveat, neither the variables in W nor those in P can be correlated with the variables in Z (See Kumbhakar and Lovell³²), otherwise, the standard errors would be artificially low, which, in turn, would inflate the t -ratio to mistakenly provide greater statistical significance..

The doubly heteroscedastic SFA regressions are estimated by the maximum likelihood method separately in the s regressions. After estimation, we use the formula due to Jondrow *et al.*³³ to obtain estimated firm-specific inefficiency for the half-normal model as follows:

$$E[u|\varepsilon] = \frac{\sigma\lambda}{1+\lambda^2} \left[\frac{\phi\varepsilon\lambda/\sigma}{1-\Phi(\varepsilon\lambda/\sigma)} - \frac{\varepsilon\lambda}{\sigma} \right], \quad (16.22)$$

where $\varepsilon = u + v$, $\lambda = \sigma_u/\sigma_v$, $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$, while $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal cumulative distribution and density functions, respectively.³⁴

$\beta_{ro} + \sum_{k=1}^K \beta_{rk} \ln Z_{kjt} + v_{rjt}$ represents the stochastic minimum environmental impacts on the r -th slack. Given the environmental impacts and random shocks, if $u_{rjt} > 0$, it captures the true effect of management incompetence on the r -th slack; otherwise, the management is regarded truly efficient in producing output r .

Data Adjustment Method We proceed to adjust the output data with a filter-out mechanism as follows:

$$\begin{aligned} y_{rjt}^a &= y_{rjt}(1 + \beta_{ro} + \sum_{k=1}^K \beta_{rk} \ln Z_{kjt} + v_{rjt}) \\ &= y_{rjt} \left(1 + \frac{s_{rjt}^{+*}}{y_{rjt}} - u_{rjt}\right) \quad (r = 1, \dots, s) \end{aligned} \quad (16.23)$$

where y_{rjt}^a is the adjusted data on output r . In words, the outputs are adjusted to an ideal level when there were absences of environmental influences and random shocks. See Fried *et al.* (2002)³⁵ and Avkiran and Rowlands (2006)³⁶ for other adjustment formulas. Tone and Tsutsui (2006)³⁷ propose a tuning method for this adjustment.

Re-Use of DEA (3rd Stage DEA) In the final stage, we feed the adjusted outputs along with the original inputs into the same SBM model to measure efficiency scores. As another caveat, since the underlying production possibility set has been replaced, the individual efficiency scores of the first and third stages are not numerically comparable. Instead, what is more informative is the tendency of the mean measured efficiency.

16.6.3 An Application to Japanese Banking

We now illustrate how the proposed multistage framework works with an empirical application to the Japanese banking industry.

Sample Selection We construct the panel data over the period of fiscal years (FY) 1997-2001 from the financial statements of individual banks. There are 660 observations to be used in this empirical application. Interested readers are referred to Liu and Tone (2006) for a detailed description of the data set.

1st-Stage DEA Analysis In this stage, we define the inputs and the outputs of credit risk management, on the basis of which we conduct the typical static DEA analysis to estimate efficiency scores and the output slacks.

Defining Inputs and Outputs Cooper *et al.*³⁸ suggest that a notable recent development in DEA literature is to employ DEA models as optimization platforms so as to expand the scope beyond conventional efficiency evaluations. For this reason, we do not confine our attention to the conventional choices of inputs and outputs. Instead, following Brockett *et al.*³⁹ (also see Chapter 6), as well as the profit-oriented approach of Berger *et al.*,⁴⁰ we choose three inputs, namely: (1) interest expenses (IE); (2) credit costs (CC); and (3) general and administrative expenses (GAE); and two outputs: (1) interest-accruing loans (IAL); and (2) lending revenues (LRV).⁴¹

According to the microeconomics of banking (Freixas and Rochet⁴²), a bank specializes in information production. It incurs expenses in screening for viable

projects, in monitoring borrowers to prevent inappropriate behaviors, in auditing a borrower in the case of default, and in establishing long-term business relationships with borrowers. Such expenses are recorded as GAE.

After underwriting the credit commitment, the bank pays interest expenses to mobilize funds from investors. For prudent or regulatory consideration, the bank has to provide credit cost against possible loan losses. Alternatively, credit costs, by definition, can be regarded as a bad outcome from lending activities. It is therefore put on the input side in the customary DEA modeling fashion. Since credit cost is normally proportional to the outstanding amount of loans, it is proper to include it in the production possibility set.

Given these inputs, a sound bank is more capable of managing higher volumes of performing loans, and earning more revenue from lending activities. Alternatively, it is more straightforward to say that, at given level of outputs, a weaker bank usually has to pay larger amounts of interest expenses, to set aside more reserves for expected loan losses, and to spend more in legal expenses to handle the bad loans. The SBM-type models discriminate efficient banks from inefficient ones in exactly the same way.

Selecting DEA Model Specification We focus on the output oriented DEA model for two reasons. The first reason is due to the institutional constraints in Japan. Given the business conditions, bankers are faced with real rigidity in the labor market, e.g., the seniority and life-long employment system. The second reason is operational: In the assets and liabilities management, bankers normally make decisions on lending activities first, and then move to funding. This means that inputs are conditional on outputs. Therefore, the output-oriented model is potentially more relevant.

We employ the assumption of variable return to scales to account for the fact that the sizes of the banks in our sample vary greatly, ranging from large internationally active banks to small local banks.

To conclude the 1st stage analysis, we use DEA-Solver Pro to estimate the slacks and the initial efficiency scores. The summary of the initial efficiency scores will be presented later in Table 16.2 for the purpose of inter-stage comparison.

2nd-Stage SFA Analysis In this stage, we select a set of environmental variables (or, prime regressors) to explain the variation in the output slacks, and another two sets of variables (or, variance regressors) for the two heteroscedastic models. We then estimate the two output slack equations separately with LIMDEP 8.0.

Prime Regressors We select four types of variables to proxy the environments in which the Japanese banks operate, i.e., macroeconomic conditions, market conditions, financial liberalizations and risk pricing practices. The macroeconomic conditions are summarized by the growth index of the gross domestic product (GDP). The market condition variables include land and stock

price indices and Japan premium. Regarding financial liberalization, we take the ratio of monetary aggregate to GDP and that of bank lending to GDP to proxy the financially innovative activities and financial depth in Japan. Lastly, we construct short and long-term average risk spreads for each of the three types of commercial banks to proxy the risk pricing practices in Japan. In order to reduce data noise, the GDP deflator (based on 1995 constant prices) was applied to all collected data denominated by the yen.

Variance Regressors Heteroscedasticity in the symmetric error component, v_{ijt} , is normally assumed to be a function of size-related variables, while heteroscedasticity in the one-sided inefficiency error component, u_{ijt} , is correlated with management-related variables. For the former set of variables, charter type dummies are used as a proxy for size. Regional banks are fairly stable during the sample period, and hence are treated as the base group in regressions, whereas city banks and Type II regional banks are modeled as the treatment groups by constructing a dummy to indicate each charter type.

For the latter set of management-related variables, three additional dummies are constructed to incorporate such important events as public fund injections, bank insolvencies, and bank take-overs that materially affected management performance.

Furthermore, two bank-specific variables, i.e. the capital adequacy ratio (CAR) and the non-performing loan ratio (NPL) are highly relevant in credit risk management and, hence, should be included as variance regressors. However, a priori, we expect some forms of association between this pair of variables and the prime regressors. Recall the caveat in Section 16.6.2, possible correlations between prime and variance regressors should be double checked and, if detected, excluded from the second stage analysis.

We utilize the fixed effects linear model for panel data to formulate the problem in a standard fashion (Greene⁴³) as follows:

$$CAR_{jt} = \alpha_j + Z'_{jt}\phi + \varepsilon_{jt}. \quad (16.24)$$

Here CAR, the Capital Adequacy Ratio, is the dependent variable; Z is the vector of prime regressors in logarithmic form; $Z'_{jt}\phi$ is the correlation between CAR and prime regressors that needs to be excluded from the SFA analysis; α_j is the groupwise fixed effects, or bank heterogeneity in CAR, and inserted into W of equation (16.20). ε_{jt} is the classical error term, which is usually considered to contain management-related information, and hence is included in P of equation (16.21).

The same treatment is also applied to the case of NPL. The two equations are estimated separately by ordinary least squares. For brevity, we omit the results of the auxiliary regressions here. See Liu and Tone (2006).

SFA Estimations The empirical results of the SFA estimation are presented in Table 16.1, which suggests that the output slacks are indeed affected by the environmental factors. The estimated coefficients on all significant variables

have the right signs. In the equation for the slacks in IAL, the estimated coefficients on the financial liberalization indicator, the real GDP growth index, the long-term risk spread, the Japan premium, and the real bankruptcy debt per case are significant with both statistical and economic significance, and consistent with *a priori* predictions.

Table 16.1. Stochastic Frontier Estimation Results

Independent Variables	Dependent Variables			
	Slacks in IAL		Slacks in LRV	
	Parameter	t-ratio	Parameter	t-ratio
Primary Index Equation for Model				
Constant	5.459*** ^a	2.632	-33.454***	-5.436
Monetary aggregate/GDP	0.042	0.869	-0.096	-0.937
Bank lending/GDP	-0.111**	-2.174	n.a. ^b	n.a.
Real GDP growth index	-1.078***	-2.594	5.626***	4.708
Short-term risk spread	-0.044	-0.189	1.657*	1.786
Long-term risk spread	0.110***	7.288	0.089	0.823
Japan premium	0.036***	3.596	-0.073***	-3.078
Real bankrupt debt/case	0.030**	1.993	0.121***	3.831
Real land price index	-0.047	-0.346	0.352*	1.725
Real stock price index	0.011	0.153	1.203***	6.065
Variance parameters for composed error				
λ	758.8***	3.659	984.1	1.319
σ_u	0.761***	7.737	2.313***	3.146
Heteroscedasticity in symmetric component (v)				
Bank heterogeneity in NPL	16.780***	20.031	16.768***	8.289
Bank heterogeneity in CAR	1.068*	1.462	2.619**	1.996
Dummy for city banks	n.a.	n.a.	0.755**	2.079
Dummy for regional banks (II)	n.a.	n.a.	-0.029	-0.188
Heteroscedasticity in one sided component (u)				
Residuals in NPL	-6.555	-0.647	-0.323	-0.024
Residuals in CAR	-5.043	-0.367	4.772	0.286
Dummy for bank bailouts	-2.704***	-6.399	-1.317*	-1.946
Dummy for taking-over	14.018***	22.731	-1.926**	-2.237
Dummy for insolvency	n.a.	n.a.	-2.675***	-2.932
Log-L	-36.227	n.a.	-524.086	n.a.
Observations	660	n.a.	660	n.a.

^a: The asterisks *, **, and *** indicate significance levels of 15%, 5%, and 1% or better, respectively.

^b: n.a. indicates non-applicable.

In the equation for the slacks in LRV, the estimated coefficients on all explanatory variables, except for the two indicators for financial liberalization, and the long-term risk spread are significant, both statistically and economically, and consistent with theoretical predictions.

Regarding the estimated variance parameters in Table 16.1, the values of λ in both equations are large in magnitude and statistically significant, indicating that the output slacks are mostly attributable to management incompetence. In the heteroscedastic models of symmetric error components, the charter dummies only work in the LRV slacks equation. They suggest city banks have significantly larger variance in random shocks than regional banks. This makes sense in that the former operate worldwide, and hence are exposed to more complicated risks, whereas the latter largely serve local customers for simpler financing needs.

In the heteroscedastic model of the one-sided component, the dummy for insolvent banks does not work well in the IAL slacks equation. The estimated coefficients on all the other dummies are significant, however, and consistent to theories in both equations.

3rd-Stage DEA Analysis After adjusting the output data using equation (16.23), we again run the same SBM model to estimate the efficiency scores. The initial and the final estimates of efficiency scores are summarized together in Table 16.2 and Figure 16.2.

As mentioned in Section 16.2.2, only the trends of change in efficiency are comparable across the stages. The mean of initially measured efficiency is volatile, and ends up losing 8.7% in FY 2001 compared to the starting level of FY 1997. Meanwhile, the distributions of the individual scores generally become more dispersed, with the standard deviation increasing from 0.092 in FY 1997 to 0.146 in FY 2001. In particular, there are two notable declines coinciding with the Asian financial crises in 1997/98 and the bursting of the IT bubble in 2001.

Table 16.2. Comparison of the Initial and Final Efficiency Scores

Stage		1997	1998	1999	2000	2001
Initial estimates	Mean	0.8768	0.8082	0.8605	0.8573	0.8005
	Std.Dev.	0.0921	0.1453	0.1018	0.1065	0.1455
Final estimates	Mean	0.7981	0.7867	0.8069	0.8157	0.8822
	Std.Dev.	0.1944	0.1791	0.1846	0.1696	0.1180

In sharp contrast, the mean efficiency score measured in the final stage exhibits a more stable upward pattern, seemingly independent of the exogenous influences. There is a notable pick-up toward the end of the sample period,

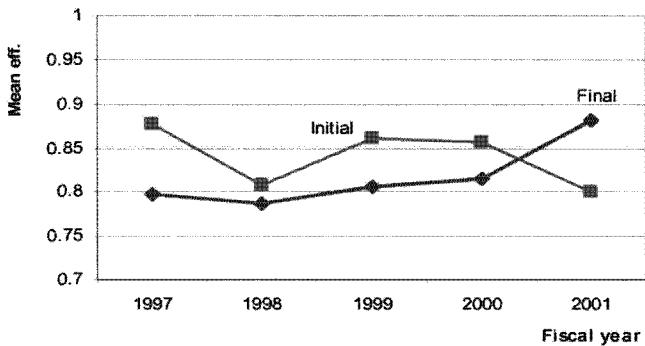


Figure 16.2. Comparison of Mean Efficiency between Initial and Final Estimates

leaving an efficiency gain of 10.54% compared to that of FY 1997. Meanwhile, the standard deviation of efficiency scores steadily narrows from 0.194 in FY 1997 to 0.118 in FY 2001. The combination of the increasing mean and the narrowing standard deviation value is consistent to the Arrow theory of learning-by-doing.

16.6.4 Discussion

In the initial stage, the typical static DEA results are affected by the influences of environmental factors and random shocks pertaining to the sample period. The development in banking productivity coincides with the stagnant economic and market conditions in Japan, with two notable declines that coincide with the Asian financial crises and the bursting of the dot.com bubble. Thus, the empirical result lends support to Greenspan,⁴⁴ who noted that risks had been repeatedly underestimated in booms, and overestimated in recessions, making the bank loan portfolio increasingly sensitive to the business cycle.

After controlling for the aforementioned impacts, the final stage efficiency estimates suggest that Japanese bankers actually kept learning and generally became more efficient over time. The result conforms well to reality. Since the late 1990s, banks have been investing heavily in IT infrastructure and special information systems to build up their capacity for risk modeling and management. Meanwhile, various standard-setting bodies and regulatory authorities have been continuously disseminating information on best practices of banking activities as well as enhancing disclosure transparency. In such a context, it is reasonable to expect bankers to be more efficient in risk management.

A recent development in Japanese banking is also worth noting. According to the Bank of Japan,⁴⁵ the bankers have increased syndicated lending to share

credit risks. It is a sharp departure from the traditional “main bank” practice. Therefore, we consider final estimates to be more convincing as they reflect events like the IT bubble which the initial estimates failed to do.

16.6.5 Summary of This Case Study

In this section, we discussed the impacts on efficiency measurement of the important assumption of deterministic data in the majority of DEA models. We then moved on to show how to relax such assumption by way of a multistage method to simultaneously allow for the impacts of data noise and environmental factors. As an illustration, we applied the framework to evaluate the performance of Japanese banks in credit risk management during the period of FY 1997-2001.

In the first stage, we found a volatile pattern in the mean efficiency score, making the multistage method legitimate. We thus employed SFA technique to control for the aforementioned impacts in the second-stage analysis, and accordingly revised the underlying production possibility set. We then estimated efficiency scores again in the final stage DEA analysis. We found the mean exhibited a more stable upward trend, while the standard deviation narrowed over time, a pattern that conforms closely to Arrow’s “theory of learning by doing.” Therefore, we conclude that the proposed method offers significant improvement on the current static DEA models in performance evaluations.

This method enables us to adjust the panel data for the environmental effects and measurement error by each cross section and to do so repeatedly. With the new dataset, we can go one step further to use Malmquist index analysis to investigate the catch-up effects and the frontier-shift effects in a more robust way. It is expected that this will invite more interesting researches along this line.

16.7 SUMMARY OF CHAPTER 16

Observed data for use in DEA may suffer from non-controllable environmental effects and statistical noise. Hence, detaching these external effects from data and uncovering the true managerial efficiency are crucial for evaluating DMUs’ performance. In this chapter, we have covered these subjects and demonstrated a case study dealing with Japanese banking using a three-stage approach.

16.8 NOTES AND SELECTED BIBLIOGRAPHY

We have listed main references to this chapter in the Notes below.

The three-stage scheme developed in the last section was successfully applied for the price-cap settings of NTT (Nippon Telegraph and Telephone) of which details can be found on the website⁴⁶ of Ministry of Internal Affairs and Communications of Japanese Government, and in Liu (2006).⁴⁷

Notes

1. A.N. Berger and D.B. Humphrey (1997), "New Approaches for Analyzing and Evaluating the Performance of Financial Institutes," *European Journal of Operational Research* 98/2, 175-212.
2. See footnote No. 1.
3. D.J. Aigner and S.F. Chu (1968), "On Estimating the Industry Production Function," *American Economic Review* 58, 826-839.
4. P. Schmidt (1976), "On the Statistical Estimation of Parametric Frontier Production Functions," *Review of Economics and Statistics* (May) 238-239.
5. W.W. Cooper (2005), "Origins, Uses of, and Relations Between Goal Programming and Data Envelopment Analysis," *Journal of Multi-Criteria Decision Analysis* 13, 1-9.
6. A. Charnes, W.W. Cooper and R.O. Ferguson (1957), "Optimal Estimation of Executive Compensation," *Management Science* I, 138-151.
7. G. Bassett and R. Koenker (1978), "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association* 73, 618-622.
8. T. Dielman and R. Pfaffenberger (1982), "LAV (Least Absolute Value) Estimation in Linear Regression: A Review," in *TIMS Studies in the Management Sciences* (Amsterdam: North Holland Publishing Co.).
9. R.D. Banker, H. Chang and W.W. Cooper (2004), "A Simulation Study of DEA and Parametric Frontier Models in the Presence of Heteroscedasticity," *European Journal of Operational Research* 153, 624-640.
10. J. Richmond (1974), "Estimating the Efficiency of Production," *International Economic Review* 15, 515-521.
11. W.H. Greene (1980), "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics* 13, 27-56.
12. J. Olson, P. Schmidt and D. Waldman (1980), "A Monte Carlo Study of Estimations of Stochastic Frontier Production Functions," *Journal of Econometrics* 18, 435-444.
13. R.D. Banker, H. Chang and W.W. Cooper (2004), "A Simulation Study of DEA and Parametric Frontier Models in the Presence of Heteroscedasticity," *European Journal of Operational Research* 153, 624-640.
14. S. Kumbhakar and C.A.K. Lovell (2000), *Stochastic Frontier Analysis* (Cambridge: Cambridge University Press).
15. J. Jondrow, C.A.K. Lovell, I.S. Materov and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model," *Journal of Econometrics* 51, 259-284.
16. V. Arnold, I. Bardhan and W.W. Cooper (1996), "New Uses of DEA and Statistical Regressions for Efficiency Evaluation and Estimation-With an Illustrative Application to Public Secondary Schools in Texas," *Annals of Operations Research* 66, 255-278.
17. I.R. Bardhan, W.W. Cooper and S.C. Kumbhakar (1998), "A Simulation Study of Joint Uses of Data Envelopment Analysis and Statistical Regressions for Production Function Estimations and Efficiency Evaluation," *Journal of Productivity Analysis* 9, 249-278
18. I.R. Bardhan, W.W. Cooper and S.C. Kumbhakar (1998), "A Simulation Study of Joint Uses of Data Envelopment Analysis and Statistical Regressions for Production Function Estimations and Efficiency Evaluation," *Journal of Productivity Analysis* 9, 249-278.
19. It is of interest to note that Bardhan *et al.* reported that the performance of SFA was greatly improved when the intercept term was augmented in the manner of COLS. This suggests that combinations of some of these regression methods might be worth exploring.
20. H.O. Fried, C.A.K. Lovell, S.S. Schmidt and S. Yaisawarng (2002), "Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis," *Journal of Productivity Analysis* 17, 157-174.
21. K.J. Arrow (1962), "The Economic Implications of Learning by Doing," *Review of Economic Studies* 29, pp.154-173.

22. International Monetary Fund. Japan - staff report for the 2002 Article IV consultation. IMF Country Report No. 02/175.
23. R.D. Banker and R.C. Morey (1986), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* 34, pp.513-21.
24. C.P. Timmer (1971), "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency," *Journal of Political Economy* 79, pp.776-94.
25. H.O. Fried, C.A.K. Lovell, S.S. Schmidt and S. Yaisawarg (2002), "Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis," *Journal of Productivity Analysis*, 17, pp.157-74.
26. The estimated optimal slacks are not unique in DEA models. However, according to the results of experiments, the admissible ranges of the optimal slacks are quite small, and this issue raises no serious problem in this context.
27. J.M. Liu and K. Tone (2006), "A Multistage Method to Measure Efficiency and its Application to Japanese Banking," *Socio-Economic Planning Sciences* (forthcoming).
28. N.K. Avkiran and T. Rowlands (2006), "How to Better Identify the True Managerial Performance: State of the Art Using DEA," *Omega, The International Journal of Management Science* (forthcoming).
29. G.E. Battese and T.J. Coelli (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," *Empirical Economics* 20, pp.325-32.
30. W.H. Greene (2002), *Econometric Modeling Guide*, Vol. 2. New York: Econometric Software Inc.
31. See Note 30.
32. S.C. Kumbhakar and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, New York: Cambridge University Press.
33. J. Jondrow, C.A.K. Lovell, I.S. Materov and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model," *Journal of Econometrics*, 19, pp.233-238.
34. The λ is a scalar in the SFA context, whereas it is an intensity vector in the usual DEA models . This chapter closely follows the notational conventions of both DEA and SFA. However, there should be no confusion since the DEA estimates on λ are not reported.
35. See the Note 25.
36. See the Note 28.
37. K. Tone and M. Tsutsui (2006), "Tuning SFA Results for Use in DEA," Working Paper, National Graduate Institute for Policy Studies.
38. W.W. Cooper, L.M. Seiford, K. Tone and J. Zhu (2006), "DEA: Past Accomplishments and Future Prospects," *Journal of Productivity Analysis*, (Forthcoming).
39. P.L. Brockett, A. Charnes, W.W. Cooper, Z. Huang and D.B. Sun (1997), "Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performance," *European Journal of Operational Research*, 98, pp.250-65.
40. A.N. Berger and L.J. Mester (2003), "Explaining the Dramatic Changes in Performance of US Banks: Technological Change, Deregulation, and Dynamic Changes in Competition," *Journal of Financial Intermediation*, 12, pp.57-95.
41. Readers are referred to Liu and Tone (see the Note 27) for the definitions and statistical summaries of the variables.
42. X. Freixas and J.C. Rochet (1997), *Microeconomics of Banking*. Cambridge, MA.: MIT Press.
43. W.H. Greene (2003), *Econometric Analysis*, 5th ed. New York: Pearson Education International.
44. A. Greenspan (2002), "Cyclicalities and banking regulation," remarks by Chairman Alan Greenspan at the Conference on Bank Structure and Competition. Federal Reserve Bank of Chicago. Chicago, Illinois.
45. Bank of Japan (2004), Overview of Japanese Banks: Observations from Financial Statements for Fiscal 2003.

46. See http://www.soumu.go.jp/s-news/2006/060411_1.html (in Japanese).
47. J. Liu (2006), "Multi-stage Use of Parametric and Non-Parametric Efficiency Measurements and its Applications," Ph.D. Thesis, National Graduate Institute for Policy Studies.

Appendix A

Linear Programming and Duality

This appendix outlines linear programming and its duality relations. Readers are referred to text books such as Gass (1985)¹, Charnes and Cooper (1961)², Mangasarian (1969)³ and Tone (1978)⁴ for details. More advanced treatments may be found in Dantzig (1963)⁵, Spivey and Thrall (1970)⁶ and Nering and Tucker (1993).⁷ Most of the discussions in this appendix are based on Tone (1978).

A.1 LINEAR PROGRAMMING AND OPTIMAL SOLUTIONS

The following problem, which minimizes a linear functional subject to a system of linear equations in nonnegative variables, is called a *linear programming problem*:

$$(P) \quad \min z = \mathbf{c}\mathbf{x} \tag{A.1}$$

$$\text{subject to } A\mathbf{x} = \mathbf{b} \tag{A.2}$$

$$\mathbf{x} \geq \mathbf{0}, \tag{A.3}$$

where $A \in R^{m \times n}$, $\mathbf{b} \in R^m$ and $\mathbf{c} \in R^n$ are given, and $\mathbf{x} \in R^n$ is the vector of variables to be determined optimally to minimize the scalar z in the objective. \mathbf{c} is a row vector and \mathbf{b} a column vector. (A.3) is called a nonnegativity constraint. Also, we assume that $m < n$ and $\text{rank}(A) = m$.

A nonnegative vector of variables \mathbf{x} that satisfies the constraints of (P) is called a *feasible solution* to the linear programming problem. A feasible solution that minimizes the objective function is called an *optimal solution*.

A.2 BASIS AND BASIC SOLUTIONS

We call a nonsingular submatrix $B \in R^{m \times m}$ of A a *basis* of A when it has the following properties: (1) it is of full rank and (2) it spans the space of solutions. We partition A into B and R and write symbolically:

$$A = [B \mid R], \tag{A.4}$$

where R is an $(m \times (n - m))$ matrix. The variable vector \mathbf{x} is similarly divided into \mathbf{x}^B and \mathbf{x}^R . \mathbf{x}^B is called *basic* and \mathbf{x}^R *nonbasic*. (A.2) can be expressed in terms of this partition as follows:

$$B\mathbf{x}^B + R\mathbf{x}^R = \mathbf{b}. \quad (\text{A.5})$$

By multiplying the above equation by B^{-1} , we have:

$$\mathbf{x}^B = B^{-1}\mathbf{b} - B^{-1}R\mathbf{x}^R. \quad (\text{A.6})$$

Thus, the basic vector \mathbf{x}^B is expressed in terms of the nonbasic vector \mathbf{x}^R . By substituting this expression into the objective function in (A.1), we have:

$$z = \mathbf{c}^B B^{-1}\mathbf{b} - (\mathbf{c}^B B^{-1}R - \mathbf{c}^R)\mathbf{x}^R. \quad (\text{A.7})$$

Now, we define a *simplex multiplier* $\boldsymbol{\pi} \in R^m$ and *simplex criterion* $\mathbf{p} \in R^{n-m}$ by

$$\boldsymbol{\pi} = \mathbf{c}^B B^{-1} \quad (\text{A.8})$$

$$\mathbf{p} = \boldsymbol{\pi}R - \mathbf{c}^R, \quad (\text{A.9})$$

where $\boldsymbol{\pi}$ and \mathbf{p} are row vectors. The following vectors are called the *basic solution* corresponding to the basis B :

$$\bar{\mathbf{x}}^B = B^{-1}\mathbf{b} \quad (\text{A.10})$$

$$\bar{\mathbf{x}}^R = \mathbf{0}. \quad (\text{A.11})$$

Obviously the basic solution is feasible for (A.2) and (A.3).

A.3 OPTIMAL BASIC SOLUTIONS

We call a basis B *optimal* if it satisfies:

$$\bar{\mathbf{x}}^B = B^{-1}\mathbf{b} \geq \mathbf{0} \quad (\text{A.12})$$

$$\mathbf{p} = \boldsymbol{\pi}R - \mathbf{c}^R \leq \mathbf{0}. \quad (\text{A.13})$$

Theorem A.1 *The basic solution corresponding to an optimal basis is the optimal solution of linear programming (P).*

Proof. It is easy to see that $(\bar{\mathbf{x}}^B = B^{-1}\mathbf{b}, \bar{\mathbf{x}}^R = \mathbf{0})$ is a feasible solution to (P). Furthermore,

$$z = \mathbf{c}^B \bar{\mathbf{x}}^B - \mathbf{p} \bar{\mathbf{x}}^R. \quad (\text{A.14})$$

Hence, by considering $\mathbf{p} \leq \mathbf{0}$, we find that z attains its minimum when $\mathbf{x}^R = \mathbf{0}$. \square

The *simplex method* for linear programming starts from a basis, reduces the objective function monotonically by changing bases and finally attains an optimal basis.

A.4 DUAL PROBLEM

Given the linear programming (P) (called the *primal problem*), there corresponds the following *dual problem* with the row vector of variables $\mathbf{y} \in R^m$.

$$(D) \quad \max w = \mathbf{yb} \quad (\text{A.15})$$

$$\text{subject to } \mathbf{yA} \leq \mathbf{c}, \quad (\text{A.16})$$

and \mathbf{y} not otherwise constrained.

Theorem A.2 *For each primal feasible solution \mathbf{x} and each dual feasible solution \mathbf{y} ,*

$$\mathbf{cx} \geq \mathbf{yb}. \quad (\text{A.17})$$

That is, the objective function value of the dual maximizing problem never exceeds that of the primal minimizing problem.

Proof. By multiplying (A.2) from the left by \mathbf{y} , we have

$$\mathbf{yAx} = \mathbf{yb}. \quad (\text{A.18})$$

By multiplying (A.16) from the right by \mathbf{x} and noting $\mathbf{x} \geq \mathbf{0}$, we have:

$$\mathbf{yAx} \leq \mathbf{cx}. \quad (\text{A.19})$$

Comparing (A.18) and (A.19),

$$\mathbf{cx} \geq \mathbf{yAx} = \mathbf{yb}. \quad (\text{A.20})$$

□

Corollary A.1 *If a primal feasible \mathbf{x}^0 and a dual feasible \mathbf{y}^0 satisfy*

$$\mathbf{cx}^0 = \mathbf{y}^0\mathbf{b}, \quad (\text{A.21})$$

then \mathbf{x}^0 is optimal for the primal and \mathbf{y}^0 is optimal for its dual.

Theorem A.3 (Duality Theorem) (i) *In a primal-dual pair of linear programs, if either the primal or the dual problem has an optimal solution, then the other does also, and the two optimal objective values are equal.*

(ii) *If either the primal or the dual problem has an unbounded solution, then the other has no feasible solution.* (iii) *If either problem has no solution then the other problem either has no solution or its solution is unbounded.*

Proof. (i) Suppose that the primal problem has an optimal solution. Then there exists an optimal basis B and $\mathbf{p} = \pi R - \mathbf{c}^R \leq \mathbf{0}$ as in (A.13). Thus,

$$\pi R \leq \mathbf{c}^R. \quad (\text{A.22})$$

However, multiplying (A.8) on the right by B ,

$$\pi B = \mathbf{c}^B. \quad (\text{A.23})$$

Hence,

$$\pi A = \pi [B|R] \leq [\mathbf{c}^B | \mathbf{c}^R] = \mathbf{c}. \quad (\text{A.24})$$

Consequently,

$$\pi A \leq \mathbf{c}. \quad (\text{A.25})$$

This shows that the simplex multiplier π for an optimal basis to the primal is feasible for the dual problem. Furthermore, it can be shown that π is optimal to the dual problem as follows: The basic solution ($\bar{x}^B = B^{-1}\mathbf{b}$, $\bar{x}^R = \mathbf{0}$) for the primal basis B has the objective value $z = \mathbf{c}^B B^{-1}\mathbf{b} = \pi\mathbf{b}$, while π has the dual objective value $w = \pi\mathbf{b}$. Hence, by Corollary A.1, π is optimal for the dual problem. Conversely, it can be demonstrated that if the dual problem has an optimal solution, then the primal problem does also and the two objective values are equal, by transforming the dual to the primal form and by observing its dual. (See Gass, *Linear Programming*, pp. 158-162, for details).

(ii) (a) If the objective function value of the primal problem is unbounded below and the dual problem has a feasible solution, then by Theorem A.2,

$$w = \mathbf{y}\mathbf{b} \leq -\infty. \quad (\text{A.26})$$

Thus, we have a contradiction. Hence, the dual has no feasible solution.

(b) On the other hand, if the objective function value of the dual problem is unbounded upward, it can be shown by similar reasoning that the primal problem is not feasible.

(iii) To demonstrate (iii), it is sufficient to show the following as an example in which both primal and dual problems have no solution.

$\text{}$	$\min \quad -x$ subject to $0 \times x = 1$ $x \geq 0$	$\text{}$	$\max \quad y$ subject to $y \times 0 \leq -1$
-----------	---	-----------	--

where x and y are scalar variables. □

A.5 SYMMETRIC DUAL PROBLEMS

The following two LPs, (P1) and (D1), are mutually dual.

$$(P1) \quad \begin{aligned} \min \quad & z = \mathbf{c}\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \quad (\text{A.27})$$

$$(D1) \quad \begin{aligned} \max \quad & w = \mathbf{y}\mathbf{b} \\ \text{subject to} \quad & \mathbf{y}\mathbf{A} \leq \mathbf{c} \\ & \mathbf{y} \geq \mathbf{0}. \end{aligned} \quad (\text{A.28}) \quad (\text{A.29})$$

The reason is that, by introducing a nonnegative slack $\lambda \in R^m$, (P1) can be rewritten as (P1') below and its dual turns out to be equivalent to (D1).

$$(P1') \quad \begin{aligned} \min z &= \mathbf{c}\mathbf{x} \\ \text{subject to } \mathbf{A}\mathbf{x} - \boldsymbol{\lambda} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{A.30}$$

This form of mutually dual problems can be depicted as Table A.1, which is expressed verbally as follows:

For the inequality \geq (\leq) constraints of the primal (dual) problem, the corresponding dual (primal) variables must be nonnegative. The constraints of the dual (primal) problem are bound to inequality \leq (\geq). The objective function is to be maximized (minimized).

This pair of LPs are called symmetric primal-dual problems. The duality theorem above holds for this pair, too.

Table A.1. Symmetric Primal-Dual Problem

$$\begin{array}{c} \boxed{x_1 \quad \dots \quad \dots \quad x_n} \quad \geq \quad 0 \\ \hline \begin{matrix} y_1 \\ \vdots \\ y_m \end{matrix} \quad \boxed{\begin{matrix} a_{11} & \dots & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ a_{m1} & \dots & \dots & a_{mn} \end{matrix}} \quad \geq \quad \boxed{\begin{matrix} b_1 \\ \vdots \\ b_m \end{matrix}} \\ \downarrow \wedge \\ \boxed{c_1 \quad \dots \quad \dots \quad c_n} \end{array}$$

A.6 COMPLEMENTARITY THEOREM

Let us transform the symmetric primal-dual problems into equality constraints by introducing nonnegative slack variables $\lambda \in R^m$ and $\mu \in R^n$, respectively.

$$(P1') \quad \begin{aligned} \min z &= \mathbf{c}\mathbf{x} \\ \text{subject to } \mathbf{A}\mathbf{x} - \boldsymbol{\lambda} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0}, \quad \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned}$$

$$(D1') \quad \begin{aligned} \max w &= \mathbf{y}\mathbf{b} \\ \text{subject to } \mathbf{y}\mathbf{A} + \boldsymbol{\mu} &= \mathbf{c} \\ \mathbf{y} &\geq \mathbf{0}, \quad \boldsymbol{\mu} \geq \mathbf{0}. \end{aligned} \tag{A.31}$$

Then, the optimality condition in Duality Theorem A.3 can be stated as follows:

Theorem A.4 (Complementarity Theorem) *Let $(\mathbf{x}, \boldsymbol{\lambda})$ and $(\mathbf{y}, \boldsymbol{\mu})$ be feasible to (P1') and (D1'), respectively. Then, $(\mathbf{x}, \boldsymbol{\lambda})$ and $(\mathbf{y}, \boldsymbol{\mu})$ are optimal*

to $(P1')$ and $(D1')$ if and only if it holds:

$$\mu x = y\lambda = 0. \quad (\text{A.33})$$

Proof.

From $\mu x = 0$, we have $(c - yA)x = 0 \Rightarrow cx = yAx$. (A.34)

From $y\lambda = 0$, we have $y(Ax - b) = 0 \Rightarrow yAx = yb$. (A.35)

Thus, $cx = yb$. By the duality theorem, x and y are optimal for the primal and the dual, respectively. \square

By (A.33), we have

$$\sum_{j=1}^n \mu_j x_j = 0, \quad \sum_{i=1}^m y_i \lambda_i = 0. \quad (\text{A.36})$$

By nonnegativity of each term in these two expressions,

$$\mu_j x_j = 0 \quad (j = 1, \dots, n) \quad (\text{A.37})$$

$$y_i \lambda_i = 0. \quad (i = 1, \dots, m) \quad (\text{A.38})$$

Thus, either the variable μ_j or the variable x_j must be zero for each j and either y_i or λ_i must be zero for each i . We called this property *complementarity*.

A.7 FARKAS' LEMMA AND THEOREM OF THE ALTERNATIVE

Theorem A.5 (Farkas' Lemma, Theorem of the Alternative) *For each $(m \times n)$ matrix A and each vector $b \in R^m$, either*

$$(I) \quad Ax = b \quad x \geq 0$$

has a solution $x \in R^n$ or

$$(II) \quad yA \leq 0 \quad yb > 0$$

has a solution $y \in R^m$ but never both.

Proof. For (I), we consider the following the primal-dual pair of LPs:

$$\begin{aligned} (P2) \quad & \min z = 0x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} (D2) \quad & \max w = yb \\ & yA \leq 0. \end{aligned}$$

If (I) has a feasible solution, then it is optimal for (P2) and hence, by the duality theorem, the optimal objective value of (D2) is 0. Therefore, (II) has no solution.

On the other hand, (D2) has a feasible solution $\mathbf{y} = \mathbf{0}$ and is not infeasible. Hence, if (P2) is infeasible, (D2) is unbounded upward. Thus, (II) has a solution. \square

A.8 STRONG THEOREM OF COMPLEMENTARITY

Theorem A.6 *For each skew-symmetric matrix K ($= -K^T$) $\in R^{n \times n}$, the inequality*

$$K\mathbf{x} \geq \mathbf{0}, \quad \mathbf{x} \geq \mathbf{0} \quad (\text{A.39})$$

has a solution $\bar{\mathbf{x}}$ such that

$$K\bar{\mathbf{x}} + \bar{\mathbf{x}} > \mathbf{0}. \quad (\text{A.40})$$

Proof. Let \mathbf{e}_j ($j = 1, \dots, n$) be the j -th unit vector and the system (P_j) be,

$$\begin{aligned} (P_j) \quad K\mathbf{x} &\geq \mathbf{0} \\ \mathbf{x} &\geq \mathbf{0}, \quad \mathbf{e}_j \mathbf{x} > 0. \end{aligned}$$

If (P_j) has a solution $\bar{\mathbf{x}}^j \in R^n$, then we have:

$$(K\bar{\mathbf{x}}^j)_j \geq 0, \quad \bar{\mathbf{x}}^j \geq \mathbf{0}, \quad \mathbf{e}_j \bar{\mathbf{x}}^j = (\bar{\mathbf{x}}^j)_j > 0$$

and hence

$$(K\bar{\mathbf{x}}^j)_j + (\bar{\mathbf{x}}^j)_j > 0.$$

If (P_j) has no solution, then by Farkas' lemma the following system has a solution $\bar{\mathbf{v}}^j \in R^n$, $\bar{\mathbf{w}}^j \in R^n$.

$$\begin{aligned} (D_j) \quad K\mathbf{v} &= \mathbf{e}_j + \mathbf{w} \\ \mathbf{v} &\geq \mathbf{0}, \quad \mathbf{w} \geq \mathbf{0}. \end{aligned}$$

This solution satisfies:

$$(K\bar{\mathbf{v}}^j)_j = 1 + (\bar{\mathbf{w}}^j)_j > 0$$

and hence

$$(K\bar{\mathbf{v}}^j)_j + (\bar{\mathbf{v}}^j)_j > 0.$$

Since, for each $j = 1, \dots, n$, either $\bar{\mathbf{x}}^j$ or $\bar{\mathbf{v}}^j$ exists, we can define a vector $\bar{\mathbf{x}}$ by summing over j . Then $\bar{\mathbf{x}}$ satisfies:

$$K\bar{\mathbf{x}} + \bar{\mathbf{x}} > \mathbf{0}.$$

\square

Let a primal-dual pair of LPs with the coefficient $A \in R^{m \times n}$, $\mathbf{b} \in R^m$ and $\mathbf{c} \in R^n$ be (P1') and (D1') in Section A.6. Suppose they have optimal

solutions $(\bar{x}, \bar{\lambda})$ for the primal and $(\bar{y}, \bar{\mu})$ for the dual, respectively. Then, by the complementarity condition in Theorem A.4, we have:

$$\bar{\mu}_j \bar{x}_j = 0 \quad (j = 1, \dots, n) \quad (\text{A.41})$$

$$\bar{y}_i \bar{\lambda}_i = 0 \quad (i = 1, \dots, m) \quad (\text{A.42})$$

However, a stronger theorem holds:

Theorem A.7 (Strong Theorem of Complementarity) *The primal-dual pair of LPs (P1') and (D1') have optimal solutions such that, in the complementarity condition (A.41) and (A.42), if one member of the pair is 0, then the other is positive.*

Proof. Observe the system:

$$\begin{aligned} Ax - rb &\geq \mathbf{0} \\ -yA + rc &\geq \mathbf{0} \\ yb - cx &\geq \mathbf{0} \\ x \geq \mathbf{0}, \quad y &\geq \mathbf{0}, \quad r \geq 0. \end{aligned}$$

We define a matrix K and a vector w by:

$$K = \begin{pmatrix} O & A & -b \\ -A^T & O & c^T \\ b^T & -c & 0 \end{pmatrix}$$

$$w^T = (y^T, x, r)^T.$$

Then, by Theorem A.6, the system

$$Kw \geq \mathbf{0}, \quad w \geq \mathbf{0}$$

has a solution $\tilde{w}^T = (\tilde{y}^T, \tilde{x}, \tilde{r})^T$ such that

$$K\tilde{w} + \tilde{w} > \mathbf{0}.$$

This results in the following inequalities:

$$A\tilde{x} - \tilde{r}b + \tilde{y}^T > \mathbf{0} \quad (\text{A.43})$$

$$-\tilde{y}A + \tilde{r}c + \tilde{x}^T > \mathbf{0} \quad (\text{A.44})$$

$$\tilde{y}b - c\tilde{x} + \tilde{r} > 0 \quad (\text{A.45})$$

We have two cases for \tilde{r} .

(i) If $\tilde{r} > 0$, we define \bar{x} and $\bar{\lambda}$ by

$$\bar{x} = \tilde{x}/\tilde{r}, \quad \bar{y} = \tilde{y}/\tilde{r} \quad (\text{A.46})$$

$$\bar{\lambda} = A\bar{x} - b, \quad \bar{\mu} = c - \bar{y}A. \quad (\text{A.47})$$

Then, $(\bar{x}, \bar{\lambda})$ is a feasible solution of $(P1')$ and $(\bar{y}, \bar{\mu})$ is a feasible solution of $(D1')$. Furthermore, $\bar{y}b \geq c\bar{x}$. Hence, these solutions are optimal for the primal-dual pair LPs. In this case, (A.43) and (A.44) result in

$$\bar{\lambda} + \bar{y} > \mathbf{0} \quad (\text{A.48})$$

$$\bar{\mu} + \bar{x} > \mathbf{0}. \quad (\text{A.49})$$

Thus, strong complementarity holds as asserted in the theorem.

(ii) If $\tilde{r} = 0$, it cannot occur that both $(P1')$ and $(D1')$ have feasible solutions. The reason is: if they have feasible solutions x^* and y^* , then

$$Ax^* \geq b, \quad x^* \geq \mathbf{0}, \quad y^*A \leq c, \quad y^* \geq \mathbf{0}. \quad (\text{A.50})$$

Hence, we have:

$$c\tilde{x} \geq y^*A\tilde{x} \geq 0 \geq \tilde{y}Ax^* \geq \tilde{y}b. \quad (\text{A.51})$$

This contradicts (A.45) in the case $\tilde{r} = 0$. Thus, the case $\tilde{r} = 0$ cannot occur. \square

A.9 LINEAR PROGRAMMING AND DUALITY IN GENERAL FORM

As a more general LP, we consider the case when there are both nonnegative variables $x^1 \in R^k$ and sign-free variables $x^2 \in R^{n-k}$ and both inequality and equality constraints are to be satisfied as follows:

$$\begin{aligned} (LP) \quad \min \quad & z = c^1x^1 + c^2x^2 \\ \text{subject to} \quad & A_{11}x^1 + A_{12}x^2 \geq b^1 \\ & A_{21}x^1 + A_{22}x^2 = b^2 \\ & x^1 \geq \mathbf{0} \\ & x^2 \text{ free}, \end{aligned} \quad (\text{A.52})$$

where $A_{11} \in R^{l \times k}$, $A_{12} \in R^{l \times (n-k)}$, $A_{21} \in R^{(m-l) \times k}$ and $A_{22} \in R^{(m-l) \times (n-k)}$. The corresponding dual problem is expressed as follows, with variables $y^1 \in R^l$ and $y^2 \in R^{m-l}$.

$$\begin{aligned} (DP) \quad \max \quad & w = y^1b^1 + y^2b^2 \\ \text{subject to} \quad & y^1A_{11} + y^2A_{21} \leq c^1 \\ & y^1A_{12} + y^2A_{22} = c^2 \\ & y^1 \geq \mathbf{0} \\ & y^2 \text{ free}. \end{aligned} \quad (\text{A.53})$$

It can be easily demonstrated that the two problems are mutually primal-dual and the duality theorem holds between them. Table A.2 depicts the general form of the duality relation of Linear Programming.

Table A.2. General Form of Duality Relation

$$\begin{array}{c}
 \begin{array}{cc|c}
 & \geq \mathbf{0} & \text{free} \\
 & \mathbf{x}^1 & \mathbf{x}^2 \\
 \hline
 \end{array} \\
 \begin{array}{c}
 \mathbf{0} \leq \begin{array}{c|c}
 \mathbf{y}^1 & \begin{array}{c|c}
 A_{11} & A_{12} \\
 \hline
 A_{21} & A_{22}
 \end{array} \\
 \hline
 \mathbf{y}^2 & \begin{array}{c|c}
 \parallel & \parallel \\
 \mathbf{c}^1 & \mathbf{c}^2
 \end{array}
 \end{array} \\
 \begin{array}{c}
 \geq \begin{array}{c|c}
 \mathbf{b}^1 & \\
 \hline
 \mathbf{b}^2
 \end{array} \\
 = \quad \quad \quad
 \end{array}
 \end{array}$$

Now, we introduce slack variables $\lambda^1 \in R^l$ and $\mu^1 \in R^k$ to (LP) and (DP) and rewrite them as (LP') and (DP') below:

$$\begin{aligned}
 (\text{LP}') \quad \min \quad & z = \mathbf{c}^1 \mathbf{x}^1 + \mathbf{c}^2 \mathbf{x}^2 \\
 \text{subject to} \quad & A_{11} \mathbf{x}^1 + A_{12} \mathbf{x}^2 - \lambda^1 = \mathbf{b}^1 \\
 & A_{21} \mathbf{x}^1 + A_{22} \mathbf{x}^2 = \mathbf{b}^2 \\
 & \mathbf{x}^1 \geq \mathbf{0} \\
 & \mathbf{x}^2 \text{ free} \\
 & \lambda^1 \geq \mathbf{0}.
 \end{aligned} \tag{A.54}$$

$$\begin{aligned}
 (\text{DP}') \quad \max \quad & w = \mathbf{y}^1 \mathbf{b}^1 + \mathbf{y}^2 \mathbf{b}^2 \\
 \text{subject to} \quad & \mathbf{y}^1 A_{11} + \mathbf{y}^2 A_{21} + \mu^1 = \mathbf{c}^1 \\
 & \mathbf{y}^1 A_{12} + \mathbf{y}^2 A_{22} = \mathbf{c}^2 \\
 & \mathbf{y}^1 \geq \mathbf{0} \\
 & \mathbf{y}^2 \text{ free} \\
 & \mu^1 \geq \mathbf{0}.
 \end{aligned} \tag{A.55}$$

We then have the following complementarity theorem:

Corollary A.2 (Complementarity Theorem in General Form)
Let $(\mathbf{x}^1, \mathbf{x}^2, \lambda^1)$ and $(\mathbf{y}^1, \mathbf{y}^2, \mu^1)$ be feasible to (LP') and (DP'), respectively. Then they are optimal to (LP') and (DP') if and only if the relation below holds.

$$\mu^1 \mathbf{x}^1 = \mathbf{y}^1 \lambda^1 = 0. \tag{A.56}$$

Also, there exist optimal solutions that satisfy the following strong complementarity.

Corollary A.3 (Strong Theorem of Complementarity) *In the optimal solutions to the primal-dual pair LPs, (LP') and (DP') , there exist ones such that, in the complementarity condition (A.56), if one of the pair is 0, then the other is positive.*

Notes

1. S.I. Gass (1985), *Linear Programming*, 5th ed., McGraw-Hill.
2. A. Charnes and W.W. Cooper (1961), *Management Models and Industrial Applications of Linear Programming*, (Volume 1 & 2), John Wiley & Sons.
3. O.L. Mangasarian (1969), *Nonlinear Programming*, McGraw-Hill.
4. K. Tone (1978), *Mathematical Programming*, (in Japanese) Asakura, Tokyo.
5. G.B. Dantzig (1963), *Linear Programming and Extensions* (Princeton: Princeton University Press).
6. W.A. Spivey and R.M. Thrall (1970), *Linear Optimization* (New York: Holt, Rinehart and Winston).
7. E.D. Nering and A.W. Tucker (1993), *Linear Programming and Related Problems* (New York: Academic Press).

Appendix B

Introduction to DEA-Solver

This is an introduction and manual for the attached DEA-Solver. There are two versions of DEA-Solver, the “Learning Version” (called **DEA-Solver-LV**, in the attached CD) and the “Professional Version” (called **DEA-Solver-PRO**: visit the DEA-Solver website at: <http://www.saitech-inc.com/> for further information). This manual serves both versions. DEA-Solver was developed by Kaoru Tone. All responsibility is attributed to Tone, but not to Cooper and Seiford in any dimension.

B.1 PLATFORM

The platform for this software is Microsoft Excel 97/2000 or later (a trademark of Microsoft Corporation).

B.2 INSTALLATION OF DEA-SOLVER

The accompanying installer will install DEA-Solver and sample problems in the attached CD-ROM to the hard disk (C:) of your PC. Click Setup.EXE in the folder “DEA-Solver” in the CD-ROM. Just follow the instruction on the screen. The folder in the hard disk is “C:\DEA-Solver” which includes the code DEA-Solver.xls and another folder “Samples.” A shortcut to DEA-Solver.xls will be automatically put on the Desktop. If you want to install “DEA-Solver” to other drive or to other folder (not to “C:\DEA-Solver”), just copy it to the disk or to the folder you designate. For the “Professional Version” an installer will automatically install “DEA-Solver”.

B.3 NOTATION OF DEA MODELS

DEA-Solver applies the following notation for describing DEA models.

<Model Name> - <I or O> - <C or V or GRS>

where I or O corresponds to “Input”- or “Output”-orientation and C, V or GRS to “Constant”, “Variable” or “General” returns to scale. For example, “AR-I-C” means the Input oriented Assurance Region model under Constant returns-to-scale assumption. In some cases, “I or O” and/or “C or V” are omitted. For example, “CCR-I” indicates the Input oriented CCR model which is naturally under constant returns-to-scale. “FDH” has no extensions. The abbreviated model names correspond to the following models,

1. CCR = Charnes-Cooper-Rhodes model (Chapters 2, 3)
2. BCC = Banker-Charnes-Cooper model (Chapters 4, 5)
3. IRS = Increasing Returns-to-Scale model (Chapter 5)
4. DRS = Decreasing Returns-to-Scale model (Chapter 5)

5. GRS = Generalized Returns-to-Scale model (Chapter 5)
6. AR = Assurance Region model (Chapter 6)
7. ARG = Assurance Region Global model (Chapter 6)
8. NCN = Non-controllable variable model (Chapter 7)
9. NDSC = Non-discretionary variable model (Chapter 7)
10. BND = Bounded variable model (Chapter 7)
11. CAT = Categorical variable model (Chapter 7)
12. SYS = Different Systems model (Chapter 7)
13. SBM = Slacks-Based Measure model (Chapter 4)
14. Weighted SBM = Weighted Slacks-Based Measure model (Chapter 4)
15. Hybrid = Hybrid model (Chapter 4)
16. Cost = Cost efficiency model (Chapter 8)
17. New-Cost = New-Cost efficiency model (Chapter 8)
18. Revenue = Revenue efficiency model (Chapter 8)
19. New-Revenue = New-Revenue efficiency model (Chapter 8)
20. Profit = Profit efficiency model (Chapter 8)
21. New-Profit = New-Profit efficiency model (Chapter 8)
22. Ratio = Ratio efficiency model (Chapter 8)
23. Bilateral = Bilateral comparison model (Chapter 7)
24. FDH = Free Disposal Hull model (Chapter 16)
25. Window = Window Analysis (Chapter 11)
26. Malmquist = Malmquist model (Chapter 11)
27. Super-efficiency = Super-efficiency model (Chapter 10)
28. Scale Elasticity = Scale Elasticity (Chapter 12)
29. Congestion = Congestion model (Chapter 12)
30. Undesirable Output = Undesirable Outputs model (Chapter 13)

B.4 INCLUDED DEA MODELS

The “Learning Version” includes all models (except Hybrid, Malmquist, Scale Elasticity, Congestion and Undesirable Outputs) and can solve problems with up to 50 DMUs; The “Professional Version” includes all models and can deal with large-scale problems within the capacity of Excel worksheet.

B.5 PREPARATION OF THE DATA FILE

The data file should be prepared in an Excel Workbook prior to execution of DEA-Solver. The formats are as follows:

B.5.1 *The CCR, BCC, IRS, DRS, GRS, SBM, Super-Efficiency, Scale Elasticity, Congestion and FDH Models*

Figure B.1 shows an example of data file for these models.

1. The first row (Row 1)

The first row (Row 1) contains Names of Problem and Input/Output Items, i.e.,

Cell A1 = Problem Name

Cell B1, C1, ... = Names of I/O items.

The heading (I) or (O), showing them as being input or output should head the names of I/O items. The items without an (I) or (O) heading will not be considered as inputs and outputs. The ordering of (I) and (O) items is arbitrary.

2. The second row and after

The second row contains the name of the first DMU and I/O values for the corresponding I/O items. This continues up to the last DMU.

3. The scope of data domain

A data set should be bordered by at least one blank column at right and at least one blank row at bottom. This is a necessity for knowing the scope of the data domain. The data set should start from the top-left cell (A1).

4. Data sheet name

A preferable sheet name is “DAT” (not “Sheet 1”). Never use names “Score”, “Rank”, “Projection”, “Weight”, “WeightedData”, “Slack”, “RTS”, “Window”, “Decomposition”, “Graph1” and “Graph2” for data sheet. These are reserved for this software.

The sample problem “Hospital(CCR)” in Figure B.1 has 12 DMUs with two inputs “(I)Doctor” and “(I)Nurse” and two outputs “(O)Outpatient” and “(O)Inpatient”. The data set is bordered by one blank column (F) and by one blank row (14). The GRS model has the constraint $L \leq \sum_{j=1}^n \lambda_j \leq U$. The values of $L(\leq 1)$ and $U(\geq 1)$ must be supplied through the Message-Box on the display by request. Defaults are $L = 0.8$ and $U = 1.2$.

As noted in 1. above, items without an (I) or (O) heading will not be considered as inputs or outputs. So, if you delete "(I)" from "(I)Nurse" to "Nurse," then "Nurse" will not be accounted for this efficiency evaluation. Thus you can add (delete) items freely to (from) inputs and outputs without changing your data set.

	A	B	C	D	E	F
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4	C	25	160	160	55	
5	D	27	168	180	72	
6	E	22	158	94	66	
7	F	55	255	230	90	
8	G	33	235	220	88	
9	H	31	206	152	80	
10	I	30	244	190	100	
11	J	50	268	250	100	
12	K	53	306	260	147	
13	L	38	284	250	120	
14						

Figure B.1. Sample.xls in Excel Sheet

B.5.2 The AR Model

Figure B.2 exhibits an example of data for the AR (Assurance Region) model. This problem has the same inputs and outputs as in Figure B.1. The constraints for the assurance region are described in rows 15 and 16 after "one blank row" at 14. This blank row is necessary for separating the data set and the assurance region constraints. These rows read as follows: the ratio of weights "(I)Doctor" vs. "(I)Nurse" is not less than 1 and not greater than 5 and that for "(O)Outpatient" vs. "(O)Inpatient" is not greater than 0.2 and not less than 0.5. Let the weights for Doctor and Nurse be $v(1)$ and $v(2)$, respectively. Then the first constraint implies

$$1 \leq v(1)/v(2) \leq 5.$$

Similarly, the second constraint means that the weights $u(1)$ (for Outpatient) and $u(2)$ (for Inpatient) satisfies the relationship

$$0.2 \leq u(1)/u(2) \leq 0.5.$$

Notice that the weights constraint can be applied between inputs and outputs as well.

	A	B	C	D	E	F
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4	C	25	160	160	55	
5	D	27	168	180	72	
6	E	22	158	94	66	
7	F	55	255	230	90	
8	G	33	235	220	88	
9	H	31	206	152	80	
10	I	30	244	190	100	
11	J	50	268	250	100	
12	K	53	306	260	147	
13	L	38	284	250	120	
14						
15	1	(I)Doctor	(I)Nurse	5		
16	0.2	(O)Outpatient	(O)Inpatient	0.5		
17						

Figure B.2. Sample-AR.xls in Excel Sheet

B.5.3 The ARG Model

Instead of restricting ratios of virtual multipliers, this model imposes bounds on the virtual input (output) relative to the total virtual input (output). For example, in the above hospital case, the virtual input of Doctor is expressed by $v(1) \times (\text{Number of}) \text{ Doctor}$ and the total virtual input is denoted by $v(1) \times (\text{Number of}) \text{ Doctor} + v(2) \times (\text{Number of}) \text{ Nurse}$, where $v(1)$ and $v(2)$ are weights to Doctor and Nurse, respectively. We impose lower and upper bounds, L and U , to the ratio of these two factors. Thus, we have constraints as expressed below.

$$L \leq \frac{v(1) \times \text{Doctor}}{v(1) \times \text{Doctor} + v(2) \times \text{Nurse}} \leq U.$$

In the Excel worksheet, we designate L and U along with the input (output) name as exhibited in Figure B.3. This means that $L = 0.5$ and $U = 0.8$ for Doctor in the above expression. See Section 6.3 in Chapter 6.

	A	B	C	D	E	F
1	Sample-ARG	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4	C	25	160	160	55	
5	D	27	168	180	72	
6	E	22	158	94	66	
7	F	55	255	230	90	
8	G	33	235	220	88	
9	H	31	206	152	80	
10	I	30	244	190	100	
11	J	50	268	250	100	
12	K	53	306	260	147	
13	L	38	284	250	120	
14						
15	0.5	(I)Doctor	0.8			
16	0.2	(I)Nurse	0.3			
17	0.2	(O)Outpatient	0.5			
18	0.4	(O)Inpatient	0.8			
19						

Figure B.3. Sample-ARG.xls in Excel Sheet

B.5.4 The NCN and NDSC Models

The non-controllable and non-discretionary models have basically the same data format as the CCR model. However, the uncontrollable inputs or outputs must have the headings (IN) or (ON), respectively. Figure B.4 exhibits the case where ‘Doctor’ is an uncontrollable (i.e., “non-discretionary” or “exogenously fixed”) input and ‘Inpatient’ is an uncontrollable output.

	A	B	C	D	E	F
1	Hospital	(IN)Doctor	(I)Nurse	(O)Outpatient	(ON)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4	C	25	160	160	55	
5	D	27	168	180	72	
6	E	22	158	94	66	
7	F	55	255	230	90	
8	G	33	235	220	88	
9	H	31	206	152	80	
10	I	30	244	190	100	
11	J	50	268	250	100	
12	K	53	306	260	147	
13	L	38	284	250	120	
14						

Figure B.4. Sample-NCN (NDSC).xls in Excel Sheet

B.5.5 The BND Model

The bounded inputs or outputs must have the headings (IB) or (OB). Their lower and upper bounds should be designated by the columns headed by (LB) and (UB), respectively. These (LB) and (UB) columns must be inserted immediately after the corresponding (IB) or (OB) column. Figure B.5 implies that ‘Doctor’ and ‘Inpatient’ are bounded variables and their lower and upper bounds are given by the columns (LB)Doc., (UB)Doc., (LB)Inpat., and (UB)Inpat, respectively.

	A	B	C	D	E	F	G	H	I
1	Hospital	(IB)Doc.	(LB)Doc.	(UB)Doc.	(I)Nurse	(O)Outpat	(OB)Inpat	(LB)Inpat.	(UB)Inpat.
2	A	20	15	22	151	100	90	80	100
3	B	19	15	23	131	150	50	45	55
4	C	25	20	25	160	160	55	50	60
5	D	27	21	27	168	180	72	70	76
6	E	22	20	25	158	94	66	60	80
7	F	55	45	56	255	230	90	80	100
8	G	33	31	36	235	220	88	80	95
9	H	31	29	33	206	152	80	70	90
10	I	30	28	31	244	190	100	90	110
11	J	50	45	50	268	250	100	90	120
12	K	53	45	54	306	260	147	130	160
13	L	38	30	40	284	250	120	110	130
14									

Figure B.5. Sample-BND.xls in Excel Sheet

B.5.6 The CAT, SYS and Bilateral Models

These models have basically the same data format as the CCR model. However, in the last column they must have an integer showing their category, system or bilateral group, as follows.

For the CAT model, the number starts from 1 (DMUs under the most difficult environment or with the most severe competition), 2 (in the second group of difficulty) and so on. It is recommended that the numbers are continuously assigned starting from 1.

For the SYS model, DMUs in the same system should have the same integer starting from 1.

For the Bilateral model, DMUs must be divided into two groups, denoted by 1 or 2.

Figure B.6 exhibits a sample data format for the CAT model.

	A	B	C	D	E	F
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	Cat.
2	A	20	151	100	90	1
3	B	19	131	150	50	2
4	C	25	160	160	55	2
5	D	27	168	180	72	2
6	E	22	158	94	66	1
7	F	55	255	230	90	1
8	G	33	235	220	88	2
9	H	31	206	152	80	1
10	I	30	244	190	100	1
11	J	50	268	250	100	2
12	K	53	306	260	147	2
13	L	38	284	250	120	2
14						

Figure B.6. Sample-CAT.xls in Excel Sheet

B.5.7 The Cost and New-Cost Models

The unit cost columns must have the heading (C) followed by the *input* name. The ordering of columns is arbitrary. If an input has no cost column, its cost is regarded as zero. Figure B.7 is a sample.

	A	B	C	D	E	F	G	H
1	Hospital	(I)Doctor	(C)Doctor	(I)Nurse	(C)Nurse	(O)Outpat.	(O)Inpat.	
2	A	20	500	151	100	100	90	
3	B	19	350	131	80	150	50	
4	C	25	450	160	90	160	55	
5	D	27	600	168	120	180	72	
6	E	22	300	158	70	94	66	
7	F	55	450	255	80	230	90	
8	G	33	500	235	100	220	88	
9	H	31	450	206	85	152	80	
10	I	30	380	244	76	190	100	
11	J	50	410	268	75	250	100	
12	K	53	440	306	80	260	147	
13	L	38	400	284	70	250	120	
14								

Figure B.7. Sample-Cost(New-Cost).xls in Excel Sheet

B.5.8 The Revenue and New-Revenue Models

The unit price columns must have the heading (P) followed by the *output* name. The ordering of columns is arbitrary. If an output has no price column, its price is regarded as zero. See Figure B.8 for an example.

	A	B	C	D	E	F	G	H
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpat.	(P)Outpat.	(O)Inpat.	(P)Inpat.	
2	A	20	151	100	550	90	2010	
3	B	19	131	150	400	50	1800	
4	C	25	160	160	480	55	2200	
5	D	27	168	180	600	72	3500	
6	E	22	158	94	400	66	3050	
7	F	55	255	230	430	90	3900	
8	G	33	235	220	540	88	3300	
9	H	31	206	152	420	80	3500	
10	I	30	244	190	350	100	2900	
11	J	50	268	250	410	100	2600	
12	K	53	306	260	540	147	2450	
13	L	38	284	250	295	120	3000	
14								

Figure B.8. Sample-Revenue(New-Revenue).xls in Excel Sheet

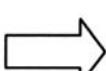
B.5.9 The Profit, New-Profit and Ratio Models

As a combination of *Cost* and *Revenue* models, these models have cost columns headed by (C) for inputs and price columns headed by (P) for outputs.

B.5.10 The Window and Malmquist Models

Figure B.9 exhibits an example of data format for Window analysis and Malmquist model. Top-left corner (A1) contains the problem name, e.g., "Car" in this example. The next right cell (B1) must include the first time period, e.g., "89." The second row beginning from the B column exhibits "I/O items", e.g., "(I)Sales" and "(O)Profit." The name of DMUs appears from the third row in the column A. The contents (observed data) follow in the third row and after. This style is repeated until the last time period. Notice that each time period is placed at the top-left corner of the corresponding frame and (I)/(O) items have the same names throughout the time period. It is not necessary to insert headings (I)/(O) to the I/O names of the second time period and after. I/O items are determined as designated in the first time period. Figure B.9 demonstrates performance of four car-manufacturers, i.e., Toyota, Nissan, Honda and Mitsubishi, during five time periods, i.e., from (19)89 to (19)93, in terms of the

input “Sales” and the output “Profit.”



A	B	C	D	E	F	G
Car	89		90		91	
DMU	(I)Sales	(O)Profit	Sales	Profit	Sales	Profit
Toyota	719	400	800	539	850	339
Nissan	358	92	401	139	418	120
Honda	264	74	275	100	280	65
Mitsubishi	190	44	203	49	231	66

H	I	J	K
92		93	
Sales	Profit	Sales	Profit
894	125	903	103
427	34	390	0
291	54	269	33
255	56	262	57

Figure B.9. Sample-Window(Malmquist).xls in Excel Sheet

B.5.11 The Hybrid Model

This model has two kinds of inputs/outputs: radial and non-radial. The radial input (output) has the heading (IR)((OR)), while the non-radial input (output)

has (INR) ((ONR)). Figure B.10 is an example.

	A	B	C	D	E	F	G	H	I
1	DMU	(IR)x1	(IR)x2	(INR)x3	(INR)x4	(OR)y1	(OR)y2	(ONR)y3	(ONR)y4
2	A	1	2	1	1	1	2	2	2
3	B	1	3	1	1	1	3	2	2
4	C	2	2	1	2	1	2	2	2
5	D	2	2	2	2	1	2	2	2
6	E	3	3	1	1	2	4	4	4
7	F	3	5	2	2	2	4	4	5

Figure B.10. Sample-Hybrid.xls in Excel Sheet

B.5.12 Weighted SBM Model

This model requires weights to inputs/outputs as data. They should be given at the rows below the main body of data set with one inserted blank row. See Figure B.11. The first column (A) has **WeightI** or **WeightO** designating input or output, respectively, and the weights to inputs or outputs follow consecutively in the order of input (output) items recorded at the top row. The values are relative, since the software normalizes them properly. Refer to (4.81)-(4.83) in Chapter 4. If they are vacant, weights are regarded as even.

	A	B	C	D	E	F
1	WSBM	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4	C	25	160	160	55	
5	D	27	168	180	72	
6	E	22	158	94	66	
7	F	55	255	230	90	
8	G	33	235	220	88	
9						
10	WeightI	10	1			
11	WeightO	1	5			
12						

Figure B.11. Sample-Weighted SBM.xls in Excel Sheet

B.5.13 The Bad Outputs Model

In addition to the ordinary (I) and (O), we employ the heading (OBad) for bad (undesirable) outputs. Figures B.12 is an example. This model requires supply of two numbers through keyboard; the total weight to good outputs and the total weight to bad outputs. Defaults are 1 and 1.

	A	B	C	D
1	Country	(I)Energy consumption	(O)GDP	(OBad)CO2
2	USA	229	104168	1529
3	China	100	12371	762
4	Russia	64	3465	392
5	Japan	51	39788	323
6	India	33	5150	292
7	Germany	33	19762	214
8	Canada	29	7157	119
9	France	27	14096	99
10	UK	22	15524	155
11	Korea	21	4767	117

Figure B.12. Sample-BadOutput.xls in Excel Sheet

B.5.14 The Non-separable Outputs Model

This model has five headings; two for input and three for output as follows:

(IS): Separable Input

(INS): Non-Separable Input

(OSGood): Separable Good Output

(ONSGood): Non-Separable Good Output

(ONSBad): Non-Separable Bad Output

Figure B.13 displays an example.

	A	B (IS)Capacity	C (INS)Fuel	D (OSGood)OtherGene	E (ONSGood)FossilGene	F (ONSBad)NOx
1	DMU					
2	1996A	129	559	178	544	156
3	1996B	86	274	102	246	54
4	1996C	92	225	289	194	57
5	1996D	105	289	243	257	117
6	1996E	83	183	53	178	52
7	1996F	115	458	48	435	105
8	1996G	153	359	265	332	110
9	1996H	200	431	480	408	163
10	2000A	141	637	149	599	123
11	2000B	86	330	109	307	48
12	2000C	92	283	304	247	65
13	2000D	121	334	274	315	69
14	2000E	78	211	50	193	42
15	2000F	122	441	82	433	95
16	2000G	164	431	277	412	94
17	2000H	200	420	569	440	100

Figure B.13. Sample-NonSeparable.xls in Excel Sheet

B.6 STARTING DEA-SOLVER

After completion of the data file in an Excel sheet on an Excel book as mentioned above, close the data file and click either the icon or the file “DEA-Solver” in Explorer. This starts DEA-Solver. First, click “Enable Macros” and then follow the instructions on the display.

Otherwise if the file “DEA-Solver” is already open (loaded), click “Tools” on the *Menu Bar*, then select “Macro” and click “Macros.” Finally, click “Run” on the Macro.

This Solver proceeds as follows,

1. Selection of a DEA model
2. Selection of a data set in Excel Worksheet
3. Selection of a Workbook for saving the results of computation and
4. DEA computation

B.7 RESULTS

The results of computation are stored in the selected Excel workbook. The following worksheets contain the results, although some models lack some of them.

1. Worksheet “Summary”

This worksheet shows statistics on data and a summary report of results obtained.

2. Worksheet “Score”

This worksheet contains the DEA-score, reference set, λ -value for each DMU in the reference set and ranking in input and in the descending order of efficiency scores. A part of a sample Worksheet “Score” is displayed in Figure B.14, where it is shown that DMUs A, B and D are efficient (Score=1) and DMU C is inefficient (Score=0.882708) with the reference set composed of B ($\lambda_B = 0.9$) and D ($\lambda_D = 0.13889$) and so on.

No.	DMU	Score	Rank	Reference set (lambda)					
1	A	1	1	A	1				
2	B	1	1	B	1				
3	C	0.8827083	8	B	0.9	D	0.13888889		
4	D	1	1	D	1				
5	E	0.7634995	12	A	0.5794409	B	5.72E-02	D	0.1526401
6	F	0.8347712	10	B	0.2	D	1.11111111		
7	G	0.9019608	7	A	0.2588235	B	1.29411765		
8	H	0.7963338	11	A	0.3866921	B	1.35E-02	D	0.6183983
9	I	0.9603922	4	A	0.6470588	B	0.83529412		
10	J	0.8706468	9	D	1.3888889				
11	K	0.955098	6	A	0.86	D	0.96666667		
12	L	0.9582043	5	A	0.6470588	B	1.23529412		

Figure B.14. A Sample Score Sheet

3. Worksheet “Rank”

This worksheet contains the ranking of DMUs in the descending order of efficiency scores.

4. Worksheet “Projection”

This worksheet contains projections of each DMU onto the efficient frontier by the chosen model.

5. Worksheet “Weight”

Optimal weights $v(i)$ and $u(i)$ for inputs and outputs are exhibited in this worksheet. $v(0)$ corresponds to the constraints $\sum_j^n \lambda_j \geq l$ and $u(0)$ to $\sum_j^n \lambda_j \leq u$. In the BCC model where $l = u = 1$ holds, $u(0)$ stands for the value of the dual variable for this constraint.

6. Worksheet “WeightedData”

This worksheet shows the optimal weighted I/O values, $x_{ij}v(i)$ and $y_{rj}u(r)$ for each DMU_j (for $j = 1, \dots, n$).

7. Worksheet “Slack”

This worksheet contains the input excesses s^- and output shortfalls s^+ for each DMU.

8. Worksheet “RTS”

In case of the BCC, AR-I-V and AR-O-V models, the returns-to-scale characteristics are recorded in this worksheet. For inefficient DMUs, returns-to-scale characteristics are those of the (input- or output-oriented) projected DMUs on the frontier.

9. Graphsheet “Graph1”

The bar chart of the DEA scores is exhibited in this graphsheet. This graph can be redesigned using the Graph functions of Excel.

10. Graphsheet “Graph2”

The bar chart of the DEA scores in the ascending order is exhibited in this graphsheet. A sample of Graph2 is exhibited in Figure B.15.

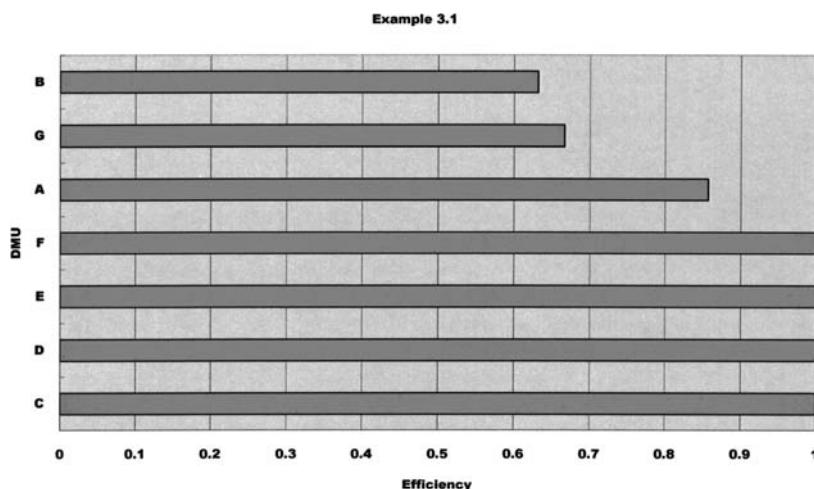


Figure B.15. A Sample Graph2

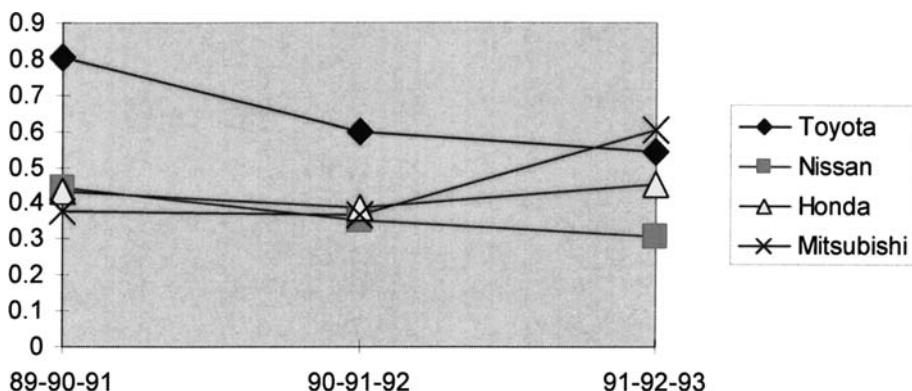
11. Worksheets “Windowk”

These sheets are only for Window models and k ranges from 1 to L (the length of time periods in the data). The contents are similar to Table 11.1 in Chapter 11. They also include two graphs, ‘Variations through Window’ and ‘Variations by Term’. We will illustrate them in the case of Sample-Window.xls in Figure B.9. Let $k = 3$ (so we deal with three adjacent years, for example). The results of computation in the case of “Window-I-C” are summarized in Table B.1.

Table B.1. Window Analysis by Three Adjacent Years

Maker	89	90	91	92	93	Average	C Average
Toyota	0.826	1	0.592			0.806	
		1	0.592	0.208		0.600	
			1	0.351	0.286	0.546	0.651
Nissan	0.381	0.515	0.426			0.441	
		0.515	0.426	0.118		0.353	
			0.720	0.200	0	0.307	0.367
Honda	0.416	0.540	0.345			0.434	
		0.540	0.345	0.275		0.387	
			0.582	0.465	0.308	0.452	0.424
Mitsubishi	0.344	0.358	0.424			0.375	
		0.358	0.424	0.326		0.369	
			0.716	0.551	0.545	0.604	0.449

From this table we can see row-wise averages of scores for each maker, which we call "Average through Window." The graph "Variations through Window" exhibits these averages. See Figure B.16.

**Figure B.16.** Variations through Window

We can also evaluate column-wise averages of scores for each maker, which we call “Average by Term.” The graph “Variations by Term” exhibits these averages. See Figure B.17.

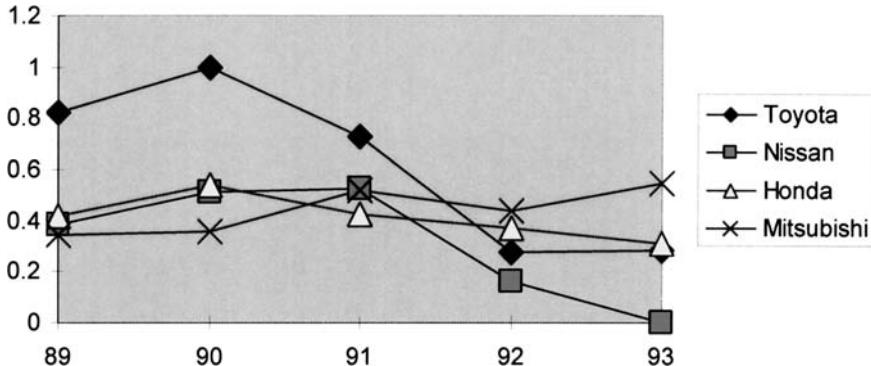


Figure B.17. Variations by Term

12. Worksheets “Malmquist k ”

These sheets are only for Malmquist models and k ranges from 1 to L (the length of time periods in the data). The worksheet “Malmquist k ” exhibits the results regarding the Malmquist index with the time interval k . For example, for the data set in Figure B.9, the worksheet “Malmquist1” contains the three indices: “Catch-up”, “Frontier-shift” and “Malmquist”, with respect to the time period pairs: (89=;90), (90=;91), (91=;92) and (92=;93). (Time interval = 1). The worksheet “Malmquist2” includes the above indices with the time period pairs: (89=;91), (90=;92) and (91=;93). (Time interval = 2). Figure B.18 shows the Malmquist index (“Malmquist1”) for the data set in Figure B.9 as evaluated by the Malmquist-I-C (input-oriented and constant returns-to-scale) model.

Malmquist	89=>90	90=>91	91=>92	92=>93	Average
Toyota	1.170124	0.512571	0.297443	0.815787	0.698981
Nissan	1.348856	0.828199	0.277361	0.001063	0.61387
Honda	1.297297	0.638393	0.799366	0.66109	0.849037
Mitsubishi	1.04232	1.183673	0.83616	1.212713	1.068717
Average	1.214649	0.790709	0.552583	0.672663	0.807651
Max	1.348856	1.183673	0.83616	1.212713	1.068717
Min	1.04232	0.512571	0.277361	0.001063	0.61387
SD	0.137262	0.29234	0.306682	0.50442	0.199356

Figure B.18. Sample Malmquist Index

This worksheet also contains the graph of the above table. See Figure B.19. The graphs for the “Catch-up” and “Frontier-shift” tables are also exhibited in each worksheet.

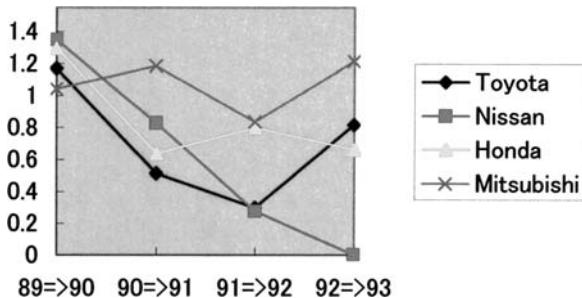


Figure B.19. Malmquist Index (Graph)

13. Worksheets “Elasticity”

The input data format for the Scale Elasticity model is the same as the CCR model, as exhibited in Table B.2 as an example.

Table B.2. Sample Data of Scale Elasticity

	(I)x	(O)y
A	2	1
B	3	3
C	4	4
D	6	5
E	7	5
F	8	4
G	9	2
H	9	1

The results of Elasticity-O (Output-oriented) model are recorded in the Worksheets “Elasticity”. (See Figure B.20.)

DMU	Scale Elasticity of Projected DMUs			
	Lower	Upper	Average	RTS
A	0.383845	8.88881	4.636327	Constant
B	0.603189	10000000	5000000	Constant
C	0.627989	1.074299	0.851144	Constant
D	0.607147	1.062706	0.834926	Constant
E	0.864194	11.76258	6.313387	Constant
F	0.69255	0.728696	0.710623	Decreasing
G	0.703994	0.731222	0.717608	Decreasing
H	0.837022	0.83702	0.837021	Decreasing
I	0.908692	0.942821	0.925757	Decreasing
J	0.282112	0.7504	0.516256	Decreasing
K	1.00E-05	0.948481	0.474246	Decreasing
L	0.101342	0.954774	0.528058	Decreasing

Figure B.20. Worksheets “Elasticity”

14. Worksheets “Congestion”

The input data format is the same with the CCR model. Figure B.21 exhibits “Congestion” and “Scale Diseconomies” regarding the data set in Table B.2. DMUs F, G and H are congested with scale diseconomies -1, -4.5 and -4.5, respectively.

DMU	Projected 1/Score	Efficiency	Scale Diseconomies
A	1	BCC-efficient	
B	1	BCC-efficient	
C	1	BCC-efficient	
D	1	BCC-efficient	
E	1	BCC-efficient	
F	1.047391	Congested	-0.775233404
G	1	BCC-efficient	
H	1	BCC-efficient	
I	1.010165	Congested	-0.271362373
J	1	BCC-efficient	
K	1	BCC-efficient	
L	1	BCC-efficient	

Figure B.21. Worksheets “Congestion”

15. Worksheets “Decomposition”

This worksheet contains the decomposition of inefficiency into inputs/outputs inefficiency. Refer to (4.84)-(4.86), (4.99)-(4.102) and (13.52)-(13.56) for details. Figure B.22 exhibits an example.

DMU	Score	Separable Capital Inefficiency	NonSeparable Fuel Inefficiency	NSBadOutput NOx Inefficiency	NSGoodOutput FossilGene Inefficiency	SGoodOutput OtherGene Inefficiency	Shortage
1996A	1	0	0	0	0	0	0
1996B	0.700458	8.68E-02	6.74E-02	0.095859	4.49E-02	6.67E-02	
1996C	1	0	0	0	0	0	0
1996D	0.621433	0.028057	0.139919	0.192154	9.33E-02	5.34E-02	
1996E	0.639287	0.209735	2.98E-02	0.103072	1.99E-02	6.67E-02	
1996F	0.890778	1.78E-02	1.10E-02	1.63E-02	7.36E-03	6.67E-02	
1996G	0.589733	0.114909	0.122869	0.143908	8.19E-02	6.67E-02	
1996H	0.714961	6.00E-02	7.12E-02	0.153358	4.75E-02	1.44E-02	
2000A	1	0	0	0	0	0	0
2000B	1	0	0	0	0	0	0
2000C	1	0	0	0	0	0	0
2000D	0.759976	0.016074	0.109058	4.80E-02	4.80E-02	5.52E-02	
2000E	0.662106	0.184805	3.51E-02	8.81E-02	2.34E-02	6.67E-02	
2000F	0.887151	1.17E-02	1.89E-02	1.33E-02	0.012625	6.67E-02	
2000G	0.693489	7.88E-02	8.60E-02	8.05E-02	5.73E-02	6.67E-02	
2000H	1	0	0	0	0	0	0

Figure B.22. Worksheets “Decomposition”

Note. The BCC, AR-I-V and AR-O-V models contain all the worksheets except “Window.” The CCR, IRS, DRS, GRS, AR-I-C, AR-O-C and SBM models contain all sheets except “RTS” and “Window.” The NCN, BND, CAT, SYS, Cost, Revenue, Profit, Ratio and FDH models produce “Summary,” (“Rank,”) “Score,” “Projection,” “Graph1” and “Graph2.” The Bilateral model shows “Summary,” “Score” and “Rank” sheets. The Window model contains only “Window” sheets.

B.8 DATA LIMITATIONS

B.8.1 Problem Size

The “Learning Version” in the attached CD can solve problems with up to 50 DMUs. In the “Professional Version,” the problem size is unlimited in terms of the number of DMUs and I/O items within the capacity of an Excel worksheet and the main memory of PC. More concretely, the data limitations for the “Professional Version” are as follows;

1. No. of DMUs must be less than 60000.
2. If $\text{No. of DMUs} \times (\text{No. of Inputs} + \text{No. of Outputs} + 2) \geq 60000$, then the “Projection” sheet will not be provided.

B.8.2 Inappropriate Data for Each Model

DMUs with the following irregular data are excluded from the comparison group as “inappropriate” DMUs. They are listed in the Worksheet “Summary.” We will adopt the following notations for this purpose.

- x_{max} (x_{min}) = the max (min) input value of the DMU concerned
- y_{max} (y_{min}) = the max (min) output value of the DMU concerned
- $cost_{max}$ ($cost_{min}$) = the max (min) unit cost of the DMU concerned
- $price_{max}$ ($price_{min}$) = the max (min) unit price of the DMU concerned

1. For the *CCR*, *BCC-I*, *IRS*, *DRS*, *GRS*, *CAT* and *SYS* models, a DMU with no positive value in inputs, i.e., $x_{max} \leq 0$, will be excluded from computation. Zero or minus values are permitted if there is at least one positive value in the inputs of the DMU concerned.
For the *BCC-O* model, DMUs with no positive value in outputs, i.e., $y_{max} \leq 0$, will be excluded from computation.
2. For the *AR* model, i.e., *AR-I-C*, *AR-I-V*, *AR-O-C* and *AR-O-V*, DMUs with $x_{min} < 0$, $x_{max} \leq 0$ or $y_{max} \leq 0$ will be excluded from the comparison group.
3. For the *FDH* model, DMUs with no positive input value, i.e., $x_{max} \leq 0$, or a negative input value, i.e., $x_{min} < 0$, will be excluded from computation.
4. For the *Cost* model, DMUs with $x_{max} \leq 0$, $x_{min} < 0$, $cost_{max} \leq 0$, or $cost_{min} < 0$ are excluded. DMUs with the current input cost ≤ 0 will also be excluded.
5. For the *Revenue*, *Profit* and *Ratio* models, DMUs with no positive input value, i.e., $x_{max} \leq 0$, no positive output value, i.e., $y_{max} \leq 0$, or with a negative output value, i.e., $y_{min} < 0$, will be excluded from computation. Furthermore, in the *Revenue* model, DMUs with $price_{max} \leq 0$, or $price_{min} < 0$ will be excluded from the comparison group. In the *Profit* model DMUs with $cost_{max} \leq 0$ or $cost_{min} < 0$ will be excluded. Finally, in the *Ratio* model, DMUs with $price_{max} \leq 0$, $price_{min} < 0$, $cost_{max} \leq 0$ or $cost_{min} < 0$ will be excluded.
6. For the *NCN* and *BND* models, negative input and output values are automatically set to zero by the program. DMUs with $x_{max} \leq 0$ in the controllable (discretionary) input variables will be excluded from the comparison group as “inappropriate” DMUs. In the *BND* model, the lower bound and the upper bound must enclose the given (observed) value, otherwise these values will be adjusted to the given value.
7. For the *Window-I-C* and *Window-I-V* models, no restriction exists for output data, i.e., positive, zero or negative values for outputs are permitted. However, DMUs with $x_{max} \leq 0$ will be characterized as being zero efficiency.

This is for purpose of completing the score matrix. So, care is needed for interpreting the results in this case. If the number of DMUs per one period (term) exceeds 255, no graph will be produced.

8. For the *SBM* model, nonpositive inputs or outputs are replaced by a small positive value.
9. For the *Bilateral* model, we cannot compare two groups if some inputs are zero for one group while the other group has all positive values for the corresponding input item.

B.9 SAMPLE PROBLEMS AND RESULTS

The attached “DEA-Solver LV (learning version)” includes the sample problems and results for all models in the folder “Samples”.

The “Professional Version” is available via <http://www.saitech-inc.com>.

B.10 SUMMARY

B.10.1 Models that Require Numbers to be Supplied through Keyboard

Table B.3 is the list of models that require supply of numbers through keyboard.

Table B.3. Models Requiring Numbers Through Keyboard

Model	Numbers	Defaults
-GRS	In the General Returns-to-Scale (GRS) models. L (lower) and U (upper) bounds.	$L = 0.8$ $U = 1.2$
BadOutput	Total weights to good outputs and bad outputs	1 and 1 (Ratio of weights =1:1)
NonSeparable	Upper bound of expansion rate of Separable Good Outputs	0.2

B.10.2 Summary of Headings to Inputs/Outputs

Table B.4 exhibits headings to input/output and samples.

Table B.4. Headings to Inputs/Outputs

Heading	Description	Example	Models employed
(I)	Input	(I)Employee	All models
(O)	Output	(O)Sales	All models
(IN)	Non-controllable or Non-discretionary input	(IN)Population	NCN (Non-controllable) NDSC (Non-discretionary)
(ON)	Non-controllable or Non-discretionary output	(ON)Area	As above
(IB)	Bounded input	(IB)Doctor	BND (Bounded variable)
(OB)	Bounded output	(OB)Attendance	As above
(LB)	Lower bound of bounded variable	(LB)Doctor	As above
(UB)	Upper bound of bounded variable	(UB)Doctor	As above
(C)	Unit cost of input	(C)Manager	Cost, New-Cost, Profit, New-Profit, Ratio
(P)	Unit price of output	(P)Laptop	Revenue, New-Revenue, Profit, New-Profit, Ratio
(OBad)	Bad (undesirable) output	(OBad)NOX	BadOutput model
(IS)	Separable input	(IS)Capital	Non-separable outputs model
(INS)	Non-Separable input	(INS)Fuel	As above
(OSGood)	Separable good output	(OSGood)Export	As above
(ONSGood)	Non-separable good output	(ONSGood)GNP	As above
(ONSBad)	Non-separable bad output	(ONSBad)CO2	As above
(IR)	Radial input	(IR)Doctor	Hybrid model
(OR)	Radial output	(OR)Inpatient	As above
(INR)	Non-radial input	(INR)Nurse	As above
(ONR)	Non-radial output	(ONR)Outpatient	As above

Appendix C Bibliography

Comprehensive bibliography of 2800 DEA references is available in the attached CD-ROM.

Author Index

- Adolphson, 240
Ahn, 95, 130, 143
Aigner, 291, 423
Ali, xxxiv, 74, 99, 117, 122, 129
Allen, 212
Andersen, 305, 309
Arnold, 66, 74, 293, 427
Aronson, 74
Arrow, 429, 439
Athanasopoulos, 206, 212
Avkiran, 430
Balk, 337
Banker, xxxi, xxxvii, 64, 87, 90, 137–138,
 142, 144, 165, 168, 172, 175, 215, 240,
 337, 363, 425, 429
Bardhan, 66, 74, 85, 113, 137, 294, 440
Barr, 212
Barua, 404
Basset, 425
Battese, 298, 432
Baumol, 279, 381
Berger, 424, 433
Bogetoft, 171
Bowlin, 118, 130
Brennan, 130
Brockett, xxx, xxxiv, 189, 193, 206, 211,
 238, 240, 255, 363, 404, 433
Bulla, 36
Caves, 323
Chang, 138, 172
Charnes, 19, 21, 33, 39, 46, 69–70, 74, 85,
 87, 90, 109, 130, 143, 165, 175–177,
 193, 216, 240, 284, 300, 303, 305–306,
 310, 315, 347, 363, 425, 443
Chen, 333, 344
Christensen, 346
Chu, 291, 423
Clark, 85
Coelli, 61, 298, 401, 432
Conrad, xxxi, xxxvii
Cook, xxxiv, 178
Cooper, 19–21, 33–36, 39, 46, 69–70, 74, 78,
 84–85, 87, 90, 109, 130, 137–138, 143,
 163, 165, 169–170, 172, 175–177, 193,
 206, 240, 254, 273, 284, 287, 291, 298,
 300, 303, 305–307, 310, 315, 347, 363,
 404, 425, 440, 443
Cummins, xxx
Dantzig, 443
Debreu, 69, 258
Deng, 223, 366, 404
Deprins, 117, 130
Desai, 206, 212
Desli, 337
Dharmapala, 35, 305
Diaz, 77
Dielman, 425
Diewert, 346
Divine, 121
Dyson, 206, 211–212
Fama, xxx, 238
Färe, 73, 112, 129, 165, 171, 258, 323, 337,
 363, 381, 392, 401
Farkas, 448
Farrell, 33, 44, 46, 69, 72, 258, 291, 306
Ferguson, 440
Ferrier, 293
Førsund, 165, 175, 363
Freixas, 433
Fried, 74, 429
Frisch, 175, 350
Fukuyama, 165, 274
Gallegos, 74
Gass, 443, 446
Gibson, 85
Gigerenzer, 304
Girod, xxxiv
Golany, xxx, 39, 85, 178, 206, 240
Gold, 401
Golden, 366

- Gong, 292
 Gonzalez-Lima, 77
 Greene, 425, 432, 435
 Greenspan, 438
 Greville, 212
 Grifell-Tatjé, 404
 Grifell-Tatje, 337
 Grosskopf, 73, 165, 172, 258, 346–347, 381, 386, 392, 404
 Haag, 305
 Halek, 39
 Haynes, 206, 212
 Hoffman, 69
 Huang, 177, 193, 206, 298, 302–303, 307
 Humphrey, 424
 Ijiri, 20
 Jagannathan, 304
 Jaska, 305
 Jensen, xxx, 238
 Johansen, 381, 401
 Jondrow, 292, 426, 432
 Kamakura, 240
 Klopp, 39, 323–324
 Koenker, 425
 Kokkelenberg, 404
 Koopmans, xxxi, 45, 68, 73
 Korostolev, 291
 Kumbhakar, 298, 306–307, 426, 432, 440
 Kwinn, 307
 Land, 299
 Langemeir, 206
 Lee, 206
 Lehman, 255
 Lelas, 303
 Levine, 240
 Lewin, 77, 85, 305
 Li, 287, 291, 298, 302–303, 307
 Lindgren, 346
 Liu, 430, 439
 Lovell, 67, 73–74, 112, 129–130, 165, 258, 291, 293, 298–299, 307, 337, 365, 392, 426, 432
 M^cCarthy, 307
 Maindiratta, 124, 165, 170
 Malmquist, 323
 Mangasarian, 443
 Mann, 233, 388
 Margaritis, 365
 Markowitz, 301
 Materov, 292, 426
 Matsui, 212
 Morey, 64, 77, 142, 170, 215, 240, 305, 429
 Nakabayashi, 415–416
 Neralic, 77, 284, 305
 Nering, 443
 Norris, 347
 Nun, 191
 Olesen, 303, 307
 Olson, 425
 Owen, 416
 Panzar, 279, 381
 Pareto, xxxi, 45, 68
 Park, xxxiv, 36, 163, 273
 Parker, 404
 Pastor, 112–113, 117, 122, 163, 273
 Perelman, 404
 Petersen, 303, 305, 307, 309
 Pfaffenbuger, 425
 Phillips, 85
 Rao, 298
 Ray, 170, 309, 337
 Rhodes, 19, 21, 33, 46, 69, 74, 240
 Richmond, 425
 Rochet, 433
 Roll, 178, 206
 Roos, 346
 Rousseau, 77, 240, 254, 305
 Rowlands, 430
 Ruefli, 121, 366, 404
 Ruiz, 112
 Russell, 129
 Saaty, 212
 Sahoo, 170, 274, 350, 357, 401
 Scheel, 376
 Schmidt, 74, 291, 425–426
 Schmitz, 39
 Seiford, 20, xxxiv, 74, 78, 85, 99, 117, 122, 124, 129–130, 171–172, 175–176, 212, 287, 291, 305, 333, 346, 376
 Semple, 240, 254, 305
 Shapley, 411
 Sharpe, 301
 Shen, 171
 Sherman, 18, 20
 Shin, 365
 Sickles, 292
 Siems, 212
 Simar, 117, 130, 306
 Simon, xxxiv, 298, 307
 Singleton, 177
 Sirvent, 112
 Smith, 177
 Spivey, 443
 Storbeck, 206, 212
 Strauss, xxxi, xxxvii
 Stutz, 124, 176
 Sueyoshi, 130, 170, 293, 326, 363
 Sugiyama, 212
 Sun, 177, 206, 326
 Svensson, 363
 Takamura, 196
 Thanassoulis, 206, 211–212
 Thomas, 39, 120

- Thompson, 35, 77, 169–170, 175, 177, 287, 365
Thore, 85, 298–299
Thrall, 35, 77, 117, 119, 122, 137–138, 162, 169–170, 175–177, 287, 306, 310, 323, 365, 443
Timmer, 429
Tone, 20, 85, 88, 100, 104, 111, 117, 128–129, 147, 160, 170, 196, 207, 212–213, 240, 255, 262, 264, 269, 271, 274, 280, 287, 333, 336, 345, 350, 357, 376, 386, 401, 415–416, 430, 433, 443
Triantis, xxxiv
Tsutsui, 269, 271, 274, 280, 376, 433
Tsybakov, 306
Tucker, 443
Tulkens, 117, 130
Valdmanis, 365, 404
Varian, 20
Waldman, 440
Walters, 85
Wang, 365
D, 171
Y, 254
Weber, 274
Wei, xxxiv, 206
Wein, 146, 279
Whinston, 404
Whitney, 233, 388
Wilcoxon, 233, 388
Willig, 279, 381
Wilson, 36, 283, 304, 306, 309
Wood, 85
Yaisawarng, 381, 386
Yamada, 212
Yu, xxxiv
Zhang, 347
Zhu, 20, 78, 171–172, 175–176, 287, 291, 305, 333, 376
Zi, xxx
Zionts, 74

Topic Index

- A priori
 - assumptions, 13
 - choices of weights, 13
 - determination, 15
 - knowledge, 177
- Absolute bounds, 178
- Academic performance, 66
- Activity, 42
- Activity analysis, 69
- ADD-efficient, 95
- Additive model, 87–88, 94, 216
 - non-discretionary variables, 216
 - vs BCC model, 96
- Admissible
 - directions, 189
 - directions
 - selecting, 189
 - nonnegative direction vectors, 186
- Advantages of DEA, 14
- Agency theory, xxx, 239
- Allocative
 - efficiency, 257–258, 272
 - efficiency
 - generalization of, 178
 - inefficiency, xxxiii, 96
 - model, 97
- AR-efficiency, 179
- AR
 - assurance region, 178, 206
 - efficiency, 179
- Archimedean property, 74
- ARG, 207
- Army recruiting, xxxiv, 324
- Assurance region, 177–178, 190, 205, 217
- Assurance region global, 207
- Assurance Region Global Model, 185
- Assurance region
 - as special case of cone-ratio, 186
 - constraints, 183, 211
 - method, 178
- vs cone-ratio, 186
- Average behavior, 4
- Average productivity, 131
- Balance sheet, 191
- Bank of Japan, 438
- Bank
 - failures, 190
 - management incompetence, 191
 - merger, 159
 - simulation, 156
 - performance, 177
- Banker-Morey model, 247
- Banks, xxx
 - Japan, 156
 - Texas, 190
- Baseball teams, 224, 253
- Basel Agreement, 190
- Basic Additive model, 94
- Basic solution, 443–444
- Basis, 443–444
- Batting average, 225
- BCC-efficient, 90, 92
- BCC-I, 119, 172
- BCC-inefficient, 92
- BCC-O, 119, 172
- BCC-projection, 92
- BCC
 - Banker, Charnes, Cooper model, 87
 - efficiency, 92
 - envelopment form, 93
 - fractional program, 91, 94
 - model, 87, 89
 - vs Additive model, 96
 - multiplier form, 91, 93
 - production possibility set, 90
 - reference set, 92
 - returns-to-scale, 136
 - two-phase procedure, 91
- Benchmark, 4
- Benchmarking, 2

- Best Linear Unbiased Estimates, 424
- Best performance, 4
- Best ratio, 13
- Best set of weights, 13
- Bilateral comparisons, 236, 240, 253
 - example, 237
 - model, 241
- Bilateral comparisons
 - model, 388
- BLUE, 424
- BND, 241
- Bootstrap method, 291
- Bounded variable model, 224
- Bounded
 - non-discretionary variable, 224
 - projection, 226
 - variable, 226
 - model, 224, 241
- Bounds
 - absolute, 178
 - lower and upper, 185
 - on multipliers, 185
 - on specific variables, 190
- Branch banks, xxix
- Branch office example, 8
- Branch store, 227
 - example, 3
- Canada, xxix
- Candidate reference point, 230
- Candidate sites, 196
- Capacity utilization, 390
 - long-run, 395
 - measure, 392
 - price-based, 392
 - short-run, 395
 - technical, 391
- CAT, 241
- Categorical
 - (classificatory) variable, 216
 - level, 227
 - algorithm, 230
 - levels expanded, 252
 - model, 228
 - variable, xxxiii, 216, 240
- Categories, 252
- CCP, 298, 303
- CCR-AR model, 179
- CCR-efficiency, 24, 46, 104
- CCR-efficient, 68, 90
- CCR-I, 70
- CCR-inefficiency, 25
- CCR-O, 70
- CCR-type models, 88
- CCR
 - Charnes, Cooper, Rhodes model, 21
 - dual problem, 41
 - model, 13, 21
- projection, 47, 49, 221
- returns-to-scale, 138
- two-phase procedure, 46
- weak efficiency form, 103
- Central tendency behavior, 4
- Chain-store example, 19
- Chance-constrained programming, xxxiv, 298, 304
- Characteristic function, 410
- Charnes-Cooper-Rousseau-Semple model, 247
- Charnes-Cooper transformation, 72, 268
- Chebychev norm, 285, 287
- Cities, xxix
- Classificatory variable, xxxiii, 216
- Coalition, 409
- Cobb-Douglas, 115, 123, 292
- Cobb-Douglas function, 424, 432
- Cobb-Douglas
 - envelopment, 124
 - form, 293
 - regression, 293
- COLS, 423
- Column views, 326
- Combustion engineering, 6
- Commercial banks, 206
- Comparing
 - DEA scores of two groups, 235
 - efficient frontiers of two groups, 236
- Comparisons
 - of different systems, 231
 - of efficiencies, 233
 - of efficiency, 241
- Complementarity theorem, 447
- Complementary condition, 46
- Complementary slackness, xxxii, 84, 184
 - and sensitivity analysis, 78
- Composed error, 291
- Computational aspects of DEA, 50
- Computational procedures, 50
- Conditional chance constraints, 303
- Cone-ratio, 177
 - as generalization of assurance region, 186
 - envelopment, 177, 190, 217
 - example, 188
 - method, 186, 205
 - formula, 187
 - vs assurance region, 186
- Confidence intervals, 291
- Congestion, 349, 353
- Congestion, 360
- Congestion, 364
 - strong, 354
 - weak, 357
- Consensus-making, 196
- Consensus formation, 196
- Consistent estimator, 291

- Constant returns-to-scale, 87, 133, 138
 - assumption, 4, 42, 68, 87
- Contestable markets, 279
- Controllable
 - categorical level, 229
 - variable, 219
- Convex
 - cone, 186
 - hull, 87
- Convexity
 - assumption, 231
 - condition, 91, 122, 150
- Cost-C(V), 274
- Cost
 - efficiency, 259, 274
- Countries, xxix
- CRS, 148
- Cyprus, xxix
- Data Envelopment Analysis, 2, 33, 52
- Data file, 456
- Data
 - imprecise, xxxvii
 - inappropriate, 474
 - limitation, 473
 - variability, 304
 - variations, 283
- DEA-Solver-LV, 454
- DEA-Solver-PRO, 454
- DEA-Solver, 17, xxxii, 42, 53, 70, 119, 172, 206, 240, 250, 274, 319, 364, 378, 454
- DEA-Solver Pro, 434
- DEA-Solver
 - AR model, 457
 - Bilateral model, 460
 - BND model, 460
 - CAT model, 460
 - Cost model, 461
 - data file, 456
 - inappropriate data, 474
 - installation, 454
 - learning version, 454, 456
 - limitations, 473
 - Malmquist model, 462
 - manual, 454
 - model notation, 454
 - NCN model, 459
 - NDSC model, 459
 - New-Cost model, 461
 - New-Profit model, 462
 - New-Revenue model, 462
 - professional version, xxxii, 454, 456
 - Profit model, 462
 - Ratio model, 462
 - results, 466
 - Revenue model, 462
 - sample problems, 475
 - starting, 466
- SYS model, 460
- website, 454
- Window model, 462
- worksheets, 466
- DEA, 2
- DEA game, 405
- DEA likelihood estimator, 291
- Decision Making Unit, 22, xxx, 33
- Decomposition
 - of efficiency, 153
 - of technical efficiency, 152, 155
- Decreasing returns-to-scale, 87, 133, 151
- Definition of SBM, 100
- Degrees of discretion, 216
- Degrees of freedom, 283–284
- Demutualizations, xxx
- Department stores, 233
- DES, 382
- Deteriorating behavior, 326
- Deterministic, 299
- Diet of Japan, 196
- Different organization forms, 238
- Dimension free, 100
- Discretionary variable, 63, 216, 324
- Distance measure, 9
- DMU, 22, 33
- Dot com companies, 401
- DRS-I, 173
- DRS-O, 173
- DRS, 148
- Dual feasible solution, 445
- Dual problem, 43, 445
- Dual simplex, 52
- Dual theorem of linear programming, 209
- Duality, 443
 - relations, xxxii
 - theorem, 35, 445, 447
- Dummy variables, 293
- Economic efficiency, 11
- Economics, xxxi
- Economies of scope, 382
- Economies of scope
 - degree, 382
- Effectiveness, 66
 - vs efficiency, 191
- Effects of a merger, 159
- Efficiency, 3, 66
 - audits, 120
 - dominance, 303
 - prices, 69
 - variance, 12
 - vs effectiveness, 191
- Efficient frontier, 3, 25, 89, 94, 139
- Elasticity, 133, 145
 - of scale, 133
- Electric power industries, 233
- England, xxix

- Envelop, 3
- Envolved, 52
- Envelopment
 - form, xxxii, 52, 93
 - map, 37
 - model, 68
- Environmental effects, 429
- Equal sample size, 255
- Euclidean
 - measure, 10
- Evaluation by experts, 191
- Excel data sheet, xxxii
- Exogenously fixed, 64
 - inputs and outputs, 240
- Expert evaluations, 191
- Extended additive model, 218
- Extreme efficient point, 289
- Facet, 139
- Farrell efficiency, 44, 46, 73
- Fast-food outlets, xxxiii
- Fast-food restaurants, 64
- FDH, 117, 120, 254
 - mixed integer program, 118
 - model, 89
- Feasible solution, 443
- Fighter Command of U.S. Air Force, 63
- Financial institutions, 156
- Fixed weights, 12, 14
 - vs. variable weights, 12
- Formula for cone-ratio method, 187
- Fractional programming, 72, 268
 - formulation for BCC, 94
 - problem, 23, 280
 - program, 24
- France, xxix
- Free disposal, 73
- Free disposal hull
 - example, 117
 - model, 89, 117
- Functional efficiency, 69
- Furnace efficiency, 6
- Fuzzy sets, xxxiv
- Gauge of scale efficiency, 142
- General hospital
 - case, 188
 - example, 181
- General returns-to-scale, 119
- Generalization of allocative and price efficiency, 178
- Generalized inverse, 212
- Generalized returns-to-scale, 151
- Global technical efficiency, 152
- Groups of DMUs, 235
 - GRS-I, 173
 - GRS-O, 173
 - GRS, 119, 151
 - Halfspaces, 135
- Heteroscedasticity, 425, 432
- Hierarchical category, 227
 - model, 241
- High-energy physics laboratory, 177
- Home runs, 225
- Hospital example, 12, 174
- Hybrid, 106, 120
 - Hybrid measure of efficiency, 117
- Hyperplane, 34, 135
- Iid, 424
- IMF, 429
- Imprecise Data, xxxvii
- Imputed (shadow) price, 184
- Inappropriate data, 474
- Income statement, 191
- Inconsistent constraints, 298
- Incorporating judgement, 177
- Increasing returns-to-scale, 87, 133, 150
- Infeasibility, 318
- Input-oriented, 115
- Input-oriented non-radial super-efficiency model, 316
- Input-oriented
 - assurance region model, 206
 - BCC model, 91
 - CCR model, 41, 58
 - SBM
 - model, 154
- Input
 - data matrix, 22
 - excess, 44
 - mix inefficiency, 101
 - numeraire, 178
 - reduction, 96
 - weights, 23
- Insurance, 194
 - companies, xxix
 - industry, 238
- Inter-group comparison, 236
- Interior point algorithm, 291
- IRS-I, 173
- IRS-O, 173
- IRS, 148
- Isoquant, 258
- Japanese banking, 433
- Japanese banks, 156
- Japanese capital, 196
- Japanese Financial Supervisory Agency, 159
- Jet Engines, 36
- Joint chance constraints, 303
- Joint constraints, 303
- K-Efficiency, 69
- Kinds of efficiency, xxxiii
- Learning version of DEA-Solver, 454
- Levels of service, 229
- Library problem, 237
- LIMDEP, 434

- Linear production games, 416
 Linear program, 23
 Linear programming, xxxii, 443
 problem, 443
 Linkage constraints, 209
 Linked-cone DEA, 209
 Linked cone, 209
 Linked constraint, 205
 Local pure technical efficiency, 152
 Log-linear, 292
 Lower and upper bounds, 184
 MAD, 425
 Maintenance activities, xxix
 Malmquist index, 328
 Management Science, xxxi
 Managerial
 efficiency, 70
 performance, 191
 Mann-Whitney statistic, 255
 Marginal productivity, 131
 Massachusetts, 18
 Max-slack solution, 45, 47, 49
 Maximizing profit, 268
 Maximizing revenue/expenses ratio, 268
 Maximum capacity, 224
 Maximum likelihood method, 432
 MCDM, xxxiv
 MDI (Minimum Discrimination Information) Statistic, xxx
 Merger, 386
 effect, 159
 Methodological bias, 293
 Minimal cost, 259
 Mix
 changes, 279
 efficiency, 154
 inefficiencies, xxxiii, 45
 inefficiency, 10
 Model selection, 114
 Monotone, 100
 Most favorable weights, 25
 Most productive scale size, 142, 144, 152
 MPSS, 142
 Moving averages, xxxv
 MPSS, 142, 144, 146, 162
 fractional program, 144
 Multicriteria decision-making, xxxiv
 Multiple interacting objectives, 239
 Multiple objective
 programming, xxxiv
 Multiple optimal solutions, 102
 Multiple reference sets, 84
 Multiplicative efficiency, 124
 Multiplicative model, 89, 123, 254
 Multiplier
 bounds, 178
 form, xxxii, 52, 91, 93
 model, 35, 68, 288
 model approach, 287
 space, 30, 34
 Multistage framework, 430
 Mutual companies, 238
 Mutual funds, xxx
 NCN-projection, 221
 NCN, 220, 240
 NDRS, 150
 NDSC, 241
 Negative slack values, 194
 New-Cost, 275
 New-Profit, 276
 New-Revenue, 275
 New capital, 196
 New capital for Japan, 196
 New insights, xxix
 NIRS, 151
 Non-Archimedean
 condition, 75
 element, 74, 138, 288
 infinitesimal, 74
 Non-controllable variable, 219–220
 model, 240
 Non-decreasing returns-to-scale, 150
 Non-discretionary, xxxiii, 240
 Non-discretionary variable
 model, 222
 Non-discretionary variable
 model, 241
 Non-discretionary
 extensions, 224
 inputs and outputs, 224
 variable, 63, 215, 248, 324
 Non-increasing returns-to-scale, 151
 Non-radial super-efficiency models, 313
 Nonconvex (staircase) production possibility set, 89
 Nonnegativity constraint, 23, 196
 Nonnegativity requirements, 195
 Nonparametric statistical tests, 234
 Nonparametric statistics, 233
 Nonunique reference set, 102
 Nonzero slack, 216
 Normalization condition, 135
 NTT, 439
 Null hypothesis, 235
 Number of DMUs, 116, 327
 Numeraire, 178
 Odense University, xxxi
 OLS, 293, 423
 Operations Research, xxxi
 Optimal
 simplex tableau, 51
 solution, 443
 weights, 25
 Ordinal relations, xxxiv

- Ordinary least squares, 293
- Organization forms, xxix
- Output-input ratio, 1
- Output-oriented, 115
- Output-oriented non-radial super-efficiency model, 316
- Output-oriented
 - assurance region model, 181, 207
 - BCC model, 93
 - CCR model, 41, 59
 - model, 58
 - scale efficiency, 154
- Output-to-input ratio, 1
- Output
 - data matrix, 22
 - expansion, 96
 - mix inefficiency, 101
 - multiplier bounds, 185
 - numeraire, 178
 - shortfall, 44
 - weights, 23
- Overall
 - efficiency, 259, 272
- Pareto-Koopmans efficiency, xxxi, 45–46
- PARN, xxxi
- Partial productivity measure, 1
- Peer (comparison) group, xxx
- Peer group, 25
- Peer set, 60
- Phase II extension, 44
- Phase II objective, 51
- Phase III, 60
- Physical variance, 11
- Piecewise linear function, 134
- Plant Capacity Utilization Measure, 392
- Point-to-hyperplane correspondence, 34
- Points at infinity, 46
- Police forces, xxix
- Polyhedral convex cone, 186
- Portfolio analysis, 301
- Positive data set assumption, 41
- Positive threshold input value, 134
- Positivity assumptions, 14
- Price-cap, 439
- Price
 - efficiency, 259
 - efficiency
 - generalization of, 178
 - inefficiencies, 96
- Pricing vector, 52
- Primal-dual correspondences, 43, 91
- Primal
 - feasible solution, 445
 - problem, 445
- Principal components analysis, 225
- Probabilistic formulation, 283, 304
- Product of Input and Output Inefficiencies, viii, 101
- Production economics, xxxi
- Production function, 131, 292
- Production possibility set, 7, 41–42, 68, 89–90, 94, 131, 135, 231, 258
- Productivity, 1, 3
- Productivity Analysis Research Network (PARN), xxxi
- Professional version of DEA-Solver, xxxii, 454, 456
- Profit-C(V), 275
- Profit-maximization mix, 260
- Profit efficiency, 260, 275
- Profit maximization, 279
- Program efficiency, 70
- Program Follow Through, 70
 - experiment, 33
- Projection
 - Additive model, 95
 - BCC, 92
 - CCR, 47
 - MPSS, 142
 - AR, 180
 - bounded, 226
 - NCN, 221
 - SBM, 102
- Proportional change, 145
- Public interest, 239
- Public libraries, 220, 227
- Public programs, 273
- Public schools, 293
- Public vs private universities, 130
- Pure technical efficiency, 153
- Purely technical inefficiency, 11, 97
- Quality-of-life, xxix
- Radial
 - (proportional) efficiency, 88
 - efficiency, 45, 88
 - measure, 9, 11
- Radius of stability, 286
- Random shocks, 431
- Rank-sum-test, 233
- Rank order statistic, xxix
- Ratio-C(V), 276
- Ratio (revenue/cost) efficiency, 276
- Ratio form, 1
- Ratio form of DEA, 70
- Ratio of ratios, 35
- Reciprocal relation, 90
- Reference set, 7, 25, 47, 102
- Refusal to play, 303
- Regions, xxix
- Regression analysis vs DEA, 4
- Regression line, 4
- Regression with dummy variable values, 295
- Regular Ultra Passum Law, 350

- Regulated Electric Cooperatives, 121
- Relative efficiency, 5
- Restricted problem, 51
- Restricted tableau, 51
- Returns-to-scale
 - BCC characterization, 136–137
 - CCR characterization, 138
 - characterization, 136
 - constant, 87
 - decreasing, 87
 - efficiency, 141
 - generalized, 151
 - identifying, 148–149
 - increasing, 87
 - inefficiencies, xxxii
 - non-decreasing, 150
 - non-increasing, 151
 - reference set characterization, 148
 - variable, 87
- Returns to scale, 131, 162, 279
- Revenue-C(V), 275
- Revenue efficiency, 260
- Revenue maximization, 275
- Revenue vs. cost ratio, 268
- Risk adjusted capital, 190
- Risk coverage, 190
- Robustness, 283
- Round robin procedure, 326
- Row views, 326
- Russell input measure, 105
- S-shaped production function, 350
- Sales environment, 227
- Sample size, 255
- Satisficing, xxxiv, 298
 - behavior, 298
- SBM-efficiency, 104
- SBM-efficient, 102
- SBM-inefficient DMU, 102
- SBM-projection, 102
- SBM, 100, 119, 154, 326
- SBM model, 430
- SBM
 - definition of, 100
- Scale economies, 360
- Scale efficiency, 142, 152
- Scale efficiency change, 337
- Scale elasticity, 349–350
- Scale Elasticity, 364
- Scale inefficiency, 96
- Schools
 - Texas, 215
- SCSC, 291
- Second stage optimization, 97
- Security banks, xxx
- Semipositive, 42, 175
 - data, 41
- Semipositivity, 68
- Sensitivity, xxxv, 130, 283, 304
- Sensitivity analysis, xxxv, 283–284, 287
- Separate efficiency frontiers, xxix
- SF, 293
- SFA, 423, 426, 430
 - doubly heteroscedastic, 431
 - estimations, 435
- Shape of the Production Possibility Set, 114
- Shapley value, 411
- Sharper discrimination among DMUs, 116
- Shortage of outpatients, 181
- Simple ratio definition, xxxi
- Simplex
 - criterion, 444
 - method, 52, 444
 - multiplier, 52, 444
- Simulation experiment, 293
- Single-output-to-single-input ratio, 2
- Site
 - selection, 177, 206
- Sites for new capital of Japan, 196
- Slack vectors, 44
- Slacks-based measure of efficiency, 88, 99, 117, 154
- Slacks, 195
- Solvency, 255
- Splitting variables, 300
- Stability, xxxv, 283
 - analysis, 304
- Staircase (or step) function, 118
- Standard normal distribution, 235
- State-mandated Excellence Standards for Texas, 66
- Statistical noise, 429
- Statistical
 - approach, 291
 - characterizations, 283
 - issues, 233
 - methods, 291
 - regression, 304, 324
 - regression model, 292
 - tests of significance, 291
- Status quo, 8
- Steady behavior, 326
- Steel industry, 279
- Stochastic Frontier Analysis, 426
- Stochastic frontier analysis, 430
- Stochastic frontier
 - regression, xxxvii
- Stochastic
 - characterizations, 283
 - efficiency, 302–303
 - frontier, 280, 293
 - regression, 291
- Stock companies, 238
- Stock vs. mutual, xxix
- Strong complementarity theorem, 46

- Strong Complementary Slackness Condition, 291
- Strong disposal, 73
- Strong theorem of complementarity, 449, 46
- Sub-additive, 410
- Summary
 - of model characteristics, 114
 - of the Basic DEA Models, 114
- Super-efficient, 385
- Super-Radial, 319
- Super-SBM, 320
- Supercolliding Super Conductor, 206
- Supermarket, 227, 233
 - example, 6, 252
- Supporting hyperplane, 135
 - as dual solution, 136
 - nonunique, 136
- Surgical units, 18
- Surplus in doctors, 181
- Symmetric primal-dual problems, 447
- SYS, 241
- Target entity, 13
- TDT measure, 35
- Teaching hospitals, 18
- Team power, 225
- Technical
 - efficiency, 11, 45, 97, 153, 258–259
 - inefficiency, 10, xxxiii, 96
 - vs mix inefficiency, 97
- Texas Assessment of Academic Skills (TAAS) Test, 66
- Texas Banking Commission, 189
- Texas banks, 190
- Texas Education Agency, 65
- Texas Public Utility Commission, 120
- Texas schools, 215
- Texas State Auditor's Office, 194
- Theorem of the alternative, 448
- Thick frontiers, 424
- Time series, 324
- Total factor productivity, 1, 14
- Transformed data, 193
- Translation invariance, 97, 116–117
- Two-phase LP problem, 44
- Two-phase procedure, 74, 91
- Two-stage analysis, xxx
- Two-stage DEA-regression approach, 293
- Two-stage solution procedure, 138
- U.S. Army Recruiting Command, 323
- Undesirable outputs model, 367
- Unequal sample size, 255
- Unit ball, 286
- Unit cost, 184
- United States, xxix
- Unitized axes, 6
- Units invariance, 5, 15
 - theorem, 24
- Units invariant, 6, 24, 61, 100, 124
- Universities, xxix
 - public vs. private, 130
- University of Warwick, xxxi
- Unstable point, 286
- USAREC, 324
- Variable returns-to-scale, 87
- Variable weights, 13–14
 - vs. fixed weights, 12
- Vector of multipliers, 51
- Virtual input, 21, 23, 33
- Virtual output, 21, 23, 33
- Visual Basic, xxxii
- Wales, xxix
- Warwick, xxxi
- Waste, 11
- Weak and strong disposal assumptions, 73
- Weak disposal, 73
- Weak efficiency, 45–46, 88, 97,
- Weak efficiency, 73
 - form for CCR model, 103
- Weather conditions, xxxiii
- Web site
 - UMass, xxxi
 - Warwick, xxxi
- Website for DEA-Solver, 454
- Welfare economics, xxxi, 68
- Welfare efficiency, xxxi
- Wilcoxon-Mann-Whitney, x, 234
 - statistic, 240
- Window analysis, xxxv, 284, 323
 - column views, 326
 - row views, 326
- Wise men, 196
- Within group comparison, 236
- Zero order decision rules, 303