Studies in Fuzziness and Soft Computing

Volume 386

Series Editor

Janusz Kacprzyk, Polish Academy of Sciences, Systems Research Institute, Warsaw, Poland

The series "Studies in Fuzziness and Soft Computing" contains publications on various topics in the area of soft computing, which include fuzzy sets, rough sets, neural networks, evolutionary computation, probabilistic and evidential reasoning, multi-valued logic, and related fields. The publications within "Studies in Fuzziness and Soft Computing" are primarily monographs and edited volumes. They cover significant recent developments in the field, both of a foundational and applicable character. An important feature of the series is its short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results. Contact the series editor by e-mail: kacprzyk@ibspan.waw.pl

Indexed by ISI, DBLP and Ulrichs, SCOPUS, Zentralblatt Math, GeoRef, Current Mathematical Publications, IngentaConnect, MetaPress and Springerlink.
The books of the series are submitted for indexing to Web of Science.

More information about this series at http://www.springer.com/series/2941

Farhad Hosseinzadeh Lotfi ·
Ali Ebrahimnejad · Mohsen Vaez-Ghasemi ·
Zohreh Moghaddas

Data Envelopment Analysis with R

Springer

Farhad Hosseinzadeh Lotfi
Department of Mathematics, Science and Research Branch
Islamic Azad University
Tehran, Iran

Ali Ebrahimnejad
Department of Mathematics, Qaemshahr Branch
Islamic Azad University
Qaemshahr, Iran

Mohsen Vaez-Ghasemi
Department of Mathematics, Rasht Branch
Islamic Azad University
Rasht, Iran

Zohreh Moghaddas
Department of Mathematics, Qazvin Branch
Islamic Azad University
Qazvin, Iran

ISSN 1434-9922 ISSN 1860-0808 (electronic)
Studies in Fuzziness and Soft Computing
ISBN 978-3-030-24276-3 ISBN 978-3-030-24277-0 (eBook)
https://doi.org/10.1007/978-3-030-24277-0

© Springer Nature Switzerland AG 2020
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In memory of our teacher,
Prof. Gholamreza Jahanshaloo
Frahad dedicated this book to his parents
Ali dedicated this book to his parents,
his wife (Zeynab) and his son (Arvin)
Mohsen and Zoherh dedicated this book
to their son, Bardia

Preface

The data envelopment analysis (DEA) technique is used to identify the strengths and weaknesses of a set of decision-maker units. In this technique for each decision-making unit based on the data analysis, relative efficiency, pattern coordinates, rankings, congestion in inputs, return on the scale, Malmquist productivity index, cost efficiency, and profit efficiency can be evaluated. For each decision-making unit (DMU) belonging to a set of observed units, it is necessary to solve an independent optimization problem in order to achieve the above objectives.

Suppose a n-member set of decision making units is evaluated. In this case,

- In order to assess the efficiency of all units, at least n independent optimization problems should be solved.
- For the ranking all units, at least n independent optimization problems should be solved.
- To calculate the input congestion for all units, at least $2n$ independent optimization problems should be solved.
- To calculate the return to the scale for all units, at least $2n$ independent optimization problems should be solved.
- To calculate the Malmquist productivity index for each of the n units, at least $4n$ independent optimization problems should be solved.
- To calculate the cost efficiency for all units, at least $2n$ independent optimization problems should be solved.
- To calculate the revenue and profit efficiencies for all units, at least n independent optimization problems should be solved.

In many real applications like evaluating companies registered in a country or a geographic area and evaluating faculties of a country, it has been observed that the number of DMUs can reach thousands of units. Problems designed to achieve these goals include linear programming, nonlinear programming, and mixed planning. In such problems, the number of constraints or variables is equal to the number of DMUs under evaluation where if the structure of the problem changes from a box to network, the number of constraints will multiply. This information indicates in order to achieve the objectives of the performance evaluation for a set of DMUS, a

huge number of optimization problem should be solved and this issue can be done by the help of strong software. It should be noted that when the number of DMUs as well as assessment periods is increased, the volume of results obtained from the optimization model will also increase. For example, if 2000 DMUs are evaluated within 12 months, then 24,000 efficiency scores will obtain after solving the corresponding models. In this way, it is necessary to provide an appropriate analysis for the management by analyzing the results. Therefore, all analysts working in the field of DEA need to know a software that can solve many of the optimization problems (linear, nonlinear, and so on) and have high capability for statistical analysis of the results derived of optimization models. One of the best software that can have good performance in solving optimization models and statistical analysis of the obtained results is R software.

Mathematical science is one of the most accurate and reliable sciences, whose precision of calculation from the origin of this science up to now has never been overlooked by anyone. But gradually due to the complexity of calculations and solving equations, manual calculations were not possible in terms of accuracy and time. However with the advent of computers, the speed of problem solving has grown steadily. Note that, the specialized problems were not solvable due to their very high time consuming and so the professional software came in. With the advent of such software, the knot of many unfinished and timely issues was opened, but learning such software turned into a problem for researchers. Today, the use of specialized and advanced software is one of the requirements of the work of the researchers in achieving the goals of the research and finding correct, quick, and accurate conclusions.

R is a mathematical and object-oriented programming language designed primarily for statistical computing and data mining. This language is in fact the S version of the open-source version. The R project was conducted by Robert Gentleman and Ross Ihaka in 1995 at the Auckland University's Department of Statistics, hence the name R. This application is licensed under the GNU General Public License (GNU GPL) of the Free Software Foundation and is available free of charge. It is currently being developed by a group of statisticians called the "Software Core Team" R page at www.r-project.org. Although R is often used for statistical computations, it can also be used for matrix computations, which is similar to software such as Octave and MATLAB. The programming language R covers a wide range of optimization of linear and nonlinear models, integer and quadratic programming as well as classical statistical tests, time series analysis, classification and clustering, and has high graphical capabilities. In the R environment, C, C ++, and FORTRAN have the ability to connect and invoke the program while the expert users can directly change the R objects by using the C code. Among the features of R can be graphical facilities for data analysis and drawing diagrams, a simple and advanced programming language including conditional expressions, loops, recursive functions, and a strong set of computing operators, arrays and matrices, software packages for data mining and analysis, storage, retrieval and manipulation of data, a multi-purpose library for analytical

operations, and data mining libraries. This software is free and open source. In this way, there is a special interest to this software among scholars.

DEA with precise and imprecise data has many applications in various fields such as agriculture, environment, companies, banks, insurance, universities, schools, manpower, hospitals, countries, mechanical engineering, service, and manufacturing centers. Thus, this technique can be used for most academic disciplines such as mathematics, industrial engineering, management, accounting, finance, agriculture, the environment as a tool for analyzing results. Consequently, this book is dedicated to the use of R codes for DEA models with crisp and fuzzy data. By describing crisp and fuzzy DEA models, it represents an outstanding reference guide helping experts to solve these models with R they must face in their different work environments.

The book is divided into five chapters. In the first, a brief introduction is given of the key concepts of DEA and fuzzy sets that are most relevant for the models that are considered in the rest of the volume. The following chapter is presented to some basic definitions about R and required commands used in writing DEA models with R. Following that, the third chapter covers basic models of DEA technique as well as R codes of the models along with numerical examples. In addition to the classic performance evaluation models in DEA, developed models have been introduced in this technique that help different aspects of analytics in performance evaluation. Therefore, in the fourth chapter, the developed DEA models are introduced briefly and then the corresponding R code for these models are provided. Finally, one of the models that is most frequently the motive for analytics is considered: the fuzzy DEA (FDEA) models. The conventional DEA models require precise input and output data, which may not always be available in real-world applications. However, in real-life problems, inputs and outputs are often imprecise which can be formulated with fuzzy data. Consequently, in the fifth chapter, the main approaches for solving FDEA models are classified and the mathematical approaches of each category are described briefly. Then, R codes for each FDEA model are provided.

To conclude, we the authors wish to express our special recognition to Prof. Janusz Kacprzyk, and Dr. Leontina Di Cecco who from the very first moment accepted our book proposal and who encouraged us continually throughout the preparation of this volume. And needless to say, the authors wish to manifest our sincere gratitude to our respective families for the support, understanding, and patience that they have shown us throughout the time that we have dedicated to preparing this book.

Tehran, Iran — Farhad Hosseinzadeh Lotfi
Qaemshahr, Iran — Ali Ebrahimnejad
Rasht, Iran — Mohsen Vaez-Ghasemi
Qazvin, Iran — Zohreh Moghaddas

Contents

1 Introduction to Data Envelopment Analysis and Fuzzy Sets 1
1.1 A Brief Review on Data Envelopment Analysis 1
1.2 Basic Definitions 4
1.3 Different Models of DEA 6
1.4 Fuzzy Set Theory 10
1.5 Conclusion 15
References 15

2 Introductions and Definitions of R 19
2.1 Preliminaries of R 19
2.2 Basic Definitions 20
2.3 Definition of Different Variables Types 23
2.4 Attributes 23
2.5 Data Storage 24
2.5.1 Vectors 24
2.5.2 Matrixes 26
2.5.3 Arrays 30
2.5.4 Lists 31
2.5.5 Data Frames 32
2.6 Mathematical Operators 33
2.7 Logical Operators 33
2.8 Use R in Calculation 34
2.9 Basic Mathematical Functions 35
2.10 If Structure 36
2.11 If Conditional 37
2.11.1 Example: Consider the Following Example 37
2.12 If Else Conditional 37
2.13 For Function 38
2.14 While Command 39
2.15 Repeat Command 40

2.16 Import and Read Data in R . . . 40
2.16.1 Data Command . . . 40
2.16.2 Scan Command . . . 41
2.16.3 Read.table Command . . . 42
2.16.4 Read.delim Command . . . 42
2.16.5 Fread Command . . . 43
2.16.6 Excel_sheets Command . . . 44
2.16.7 Read_excel Command . . . 44
2.17 Storage and Writing Data . . . 44
2.17.1 Write.table Command . . . 44
2.17.2 Sink Command . . . 44
2.18 Write Functions in R . . . 45
2.19 Convert Objects . . . 45
2.20 Conclusion . . . 50
References . . . 51

3 Basic DEA Models with R Codes . . . 53
3.1 Introduction . . . 53
3.2 Input-Oriented DEA Models with R Codes . . . 54
3.2.1 Input-Oriented CCR Envelopment Model with R Code . . . 54
3.2.2 Input-Oriented CCR Multiplier Model with R Code . . . 56
3.2.3 Input-Oriented BCC Multiplier Model with R Code . . . 60
3.2.4 Input-Oriented BCC Envelopment Model with R Code . . . 63
3.3 Output-Oriented DEA Models with R Codes . . . 67
3.3.1 Output-Oriented CCR Envelopment Model with R Code . . . 67
3.3.2 Output-Oriented CCR Multiplier Model with R Code . . . 69
3.3.3 Output-Oriented BCC Multiplier Model with R Code . . . 74
3.3.4 Output-Oriented BCC Envelopment Model with R Code . . . 76
3.4 Additive DEA Models with R Codes . . . 79
3.4.1 Additive CCR Model with R Code . . . 79
3.4.2 Additive BCC Model with R Code . . . 81
3.5 R Codes for Input-Oriented DEA Multiplier Model with ε . . . 86
3.5.1 R Code for Input-Oriented BCC Multiplier Model with ε . . . 86
3.5.2 R Code for Input-Oriented CCR Multiplier Model with ε . . . 89

3.6 Two-Phase Input-Oriented DEA Envelopment Model with R Code . . . 91
3.6.1 Two-Phase Input-Oriented BCC Envelopment Model with R Code . . . 91
3.6.2 Two-Phase Input-Oriented CCR Envelopment Model with R Code . . . 94
3.7 Conclusion . . . 98
References . . . 98

4 Advanced DEA Models with R Codes . . . 99
4.1 Introduction . . . 99
4.2 AP Models with R Codes . . . 99
4.2.1 Input-Oriented AP Envelopment Model with R Code . . . 100
4.2.2 Output-Oriented AP Enveloping Model . . . 102
4.2.3 Input-Oriented AP Multiplier Model . . . 104
4.2.4 Output-Oriented AP Multiplier Model . . . 106
4.3 MAJ Super-Efficiency Model with R Code . . . 108
4.4 Norm L_1 Super-Efficiency Model with R Code . . . 111
4.5 Returns to Scale—CCR Models with R Codes . . . 114
4.5.1 Returns to Scale—CCR Envelopment Model with R Code . . . 114
4.5.2 Returns to Scale—DEA Multiplier Model with R Code . . . 118
4.6 Cost Efficiency Model with R Code . . . 122
4.7 Revenue Efficiency DEA Model with R Code . . . 124
4.8 Malmquist Productivity Index—CCR Model with R Codes . . . 127
4.8.1 Malmquist Productivity Index—CCR Multiplier Model with R Code . . . 127
4.8.2 Malmquist Productivity Index—CCR Envelopment Model with R Code . . . 131
4.9 SBM Models with R Codes . . . 136
4.9.1 First Model of SBM with R Code . . . 136
4.9.2 Second Model of SBM with R Code . . . 139
4.10 Series Network DEA Model with R Code . . . 141
4.11 Profit Efficiency DEA Model with R Code . . . 143
4.12 Modified Slack Based DEA Models with R Codes . . . 147
4.12.1 Input-Oriented Slack Based DEA Model with R Code . . . 147
4.12.2 Output-Oriented Slack Based DEA Model with R Code . . . 150

4.13 Congestion DEA Model with R Code 152
4.14 Common Set of Weights DEA Model with R Code 156
4.15 Directional Efficiency DEA Model with R Code 158
4.16 Conclusion 162
References 162

5 Fuzzy Data Envelopment Analysis Models with R Codes 163
5.1 Introduction 163
5.2 The α–Level Approach 165
5.2.1 Kao and Liu's Approach 166
5.2.2 Saati et al.'s Approach 173
5.3 The Fuzzy Ranking Approach 176
5.3.1 Guo and Tanaka's Approach 176
5.3.2 Leon et al.'s Approach 180
5.3.3 Soleimani-damaneh et al.'s Approach 188
5.4 The Possibility Approach 190
5.5 The Fuzzy Arithmetic Approach 196
5.5.1 Wang et al.'s Approach 196
5.5.2 Bhardwaj et al.'s Approach 207
5.5.3 Azar et al.'s Approach 214
5.5.4 The MOLP Approach 218
5.6 Conclusion 235
References 235