BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

Doctoral Studies Council

Doctoral School of Economic Cybernetics and Statistics

NON-PARAMETRIC METHODS AND TECHNIQUES FOR ESTIMATING EFFICIENCY IN THE EDUCATIONAL SYSTEM

Simona – Elena D. GRĂDINARU

PhD Supervisor: Prof. univ. dr. Crișan ALBU

Bucharest, 2024

Abstract

In contemporary society, while technology becomes indispensable and increasingly complex notions gain momentum, education becomes more important than ever before. In the race towards a future fueled by artificial intelligence, study programs in the fields of STEM (Science, Technology, Engineering and Mathematics) are becoming fundamental.

These disciplines provide the knowledge, tools, and skills necessary not only for understanding new technologies, but also for developing and implementing such solutions, significantly contributing to harnessing this vast potential. The competencies acquired within these study programs make their graduates highly attractive in various fields, often attracting professions that ensure the use of state-of-the-art technologies or even contribute to their development.

Experts estimate that the labor market of 2030 will have more than 80% of jobs currently unknown, not yet invented. One of the increasingly relevant missions of educational institutions in this context is to anticipate the skills needed in the future to prepare the population for such roles.

This paper represents a comprehensive research on nonparametric techniques applied in evaluating the efficiency of universities regarding the STEM study programs offered. The empirical study focuses on estimating efficiency in higher education institutions within European countries with a very high Human Development Index, corresponding to the year 2020, based on the most recent data available.

The analysis deals with and applies a series of methods for detecting extreme values necessary for obtaining samples of comparable institutions, alongside nonparametric techniques such as Data Envelopment Analysis (DEA) and Full Disposal Hull (FDH) and their more robust versions, partial order “alpha” frontiers. Furthermore, the choice of the appropriate estimator is supported by identifying assumptions about the production set, convexity and the nature of scale efficiencies, through a series of statistical tests recently introduced in the scientific literature.

Four efficiency models have been constructed to evaluate each level of higher education, bachelor’s, master’s, and doctoral, alongside accessing European programs and funds for collaboration and research in the fields of sciences. The variables chosen to describe the input and output space corresponding to each of the four selected dimensions are representative of the way universities operate from each analyzed perspective.

The data analysis is based on a wide range of methods, including some unsupervised machine learning algorithms such as Isolation Forest and DBSCAN, for which there were no previous studies applied in efficiency analysis.

A series of characteristics are detailed with regards to both performing and inefficient institutions from the four perspectives considered in the analysis. Furthermore, the efficiency analysis of the European educational system exposes a series of comparative results obtained, along with interpretations regarding the association between the efficiency of STEM study programs in higher education and the quality of life, as well as the standard of living in the analyzed European countries, through the prism of the Human Development Index.

Keywords: efficiency in education, nonparametric methods, Data Envelopment Analysis, DEA, Full-Disposal Hull, FDH, order-alpha partial frontiers, unsupervised machine learning, outlier detection, Isolation Forest, DBSCAN, dimension reduction.

CONTENTS

1. INTRODUCTORY NOTE 10

2. SCIENTIFIC LITERATURE 18

2.1. FOUNDATIONAL CONCEPTS OF EFFICIENCY 18

2.2. PARAMETRIC AND NON-PARAMETRIC METHODS 20

2.3. EVOLUTION OF NON-PARAMETRIC METHODS FOR EFFICIENCY ANALYSIS 22

2.4. THE IMPORTANCE OF STUDYING EFFICIENCY IN EDUCATION 30

2.5. CURRENT STATE OF RESEARCH IN THE EDUCATIONAL SYSTEM 33

2.6. CHALLENGES IN ESTIMATING EFFICIENCY OF UNIVERSITIES 38

2.6.1. Choosing the variables that characterize the production process 38

2.6.2. Considering quality 40

2.6.3. Economies of scale 42

2.6.4. Comparable institutions and extreme values 43

3. RESEARCH METHODOLOGY 45

3.1. SCOPE OF RESEARCH 45

3.2. FOUNDATIONAL CONCEPTS OF EFFICIENCY ANALYSIS 46

3.3. DATA ENVELOPMENT ANALYSIS (DEA) 49

3.4. FREE DISPOSAL HULL (FDH) 51

3.5. PROBABILISTIC APPROACH FOR EFFICIENCY ESTIMATORS 53

3.6. CONSISTENCY OF ESTIMATORS 54

3.7. BOOTSTRAP RESAMPLING TECHNIQUE 55

3.8. TESTING THE HYPOTHESES OF EFFICIENCY ESTIMATORS 57

3.8.1. Returns to scale 58

3.8.2. Convexity of production set 61

3.9.ADVANCED NON-PARAMETRIC METHODS FOR EFFICIENCY ANALYSIS 65

3.9.1. Order-*m* partial frontiers 66

3.9.2. Order-*alfa* partial frontiers 69

3.10. SOFTWARE TOOLS USED IN EFFICIENCY ANALYSIS 71

4. EFFICIENCY OF EUROPEAN HIGHER EDUCATION SYSTEM 73

4.1. CURRENT CONTEXT 73

4.2. EDUCATION AS A PRODUCTION PROCESS 81

4.3. DESCRIPTION OF CHOSEN DATA AND VARIABLES 82

4.4. OUTLIER DETECTION 86

4.4.1. Outlier detection with statistical criteria 87

4.4.2. Outlier detection with Machine Learning (ML) 88

4.5. PRELIMINARY ANALYSES OF UNIVERSITY PROFILES 93

4.6. EXPLORING UNIVERSITIES USING PRINCIPAL COMPONENT ANALYSIS AND DIMENSIONALITY REDUCTION 100

4.7. ASSESSING EFFICIENCY OF EUROPEAN UNIVERSITIES THROUGH THE PERSPECTIVE OF STEM STUDY PROGRAMS 103

4.7.1. Testing hypotheses for choosing the efficiency estimator 104

4.7.2. Assessing efficiency of universities in the homogenous group 106

4.7.3. Characteristics of efficient versus inefficient universities 113

4.7.4. Assessing efficiency of universities in the heterogenous group 122

4.8. COMPARATIVE RESULTS 128

4.8.1. Model 1 – Bachelor's 128

4.8.2. Model 2 – Master's 131

4.8.3. Model 3 – Doctoral 134

4.8.4. Model 4 – Participation in EU-FP projects for scientific reserch 139

4.8.5. Comparative analysis and synthesis of results 142

5. CONCLUDING REMARKS, CONTRIBUTIONS AND FUTURE RESEARCH 146

6. REFERENCES AND BIBLIOGRAPHY 154

APPENDICES 162

LIST OF ACRONYMS AND ABBREVIATIONS 179

LIST OF TABLES 180

LIST OF FIGURES 181

LIST OF APPENDICES 183