DOCTORAATSPROEFSCHRIFT

2012 | Faculteit Bedrijfseconomische Wetenschappen

Inter-national Benchmarking of Road Safety Performance and Development using Indicators and Indexes

Data Envelopment Analysis based Approaches

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Verkeerskunde, te verdedigen door:

Yongjun SHEN

Promotor: prof. dr. Geert Wets
Copromotoren: prof. dr. Tom Brijs
prof. dr. Elke Hermans

D/2012/2451/20

universiteit hasselt

Acknowledgements

At the final stage of this dissertation, which also means the end of my 4-year doctoral study, I would like to acknowledge many people for their kind help along the way.

First and foremost, my special gratitude is to prof. dr. Da Ruan, who left us last year, so suddenly, without any portent. As a respectable supervisor, an admirable senior, and a close friend, I learned a lot from him, not only his rigorous scholarship, but also his enthusiasm, responsibility, and dedication. All experiences with him were impressive and memorable. Although he will be absent from my PhD defence, I know he will be in my heart, forever.

Next, I would like to express my warmest gratitude to my promotor, prof. dr. Geert Wets, for giving me this great opportunity to work on this interesting research area. I consider myself fortunate to work under his direction and benefit from his broad knowledge and international experience.

My deepest gratitude also goes to prof. dr. Elke Hermans, my co-promotor, who has offered not only meticulous supervision and great patience during my whole PhD period but also warm friendship and constant support. I am afraid that my English, even my Chinese, would not be still enough to find right words to express my gratefulness.

My sincere appreciation further goes to my co-promotor, prof. dr. Tom Brijs, and the member of my PhD commission, prof. dr. Koen Vanhoof, for their insightful guidance throughout this process that has encouraged me to have a better understanding of my research.

Moreover, I am indebted to the external members of the jury, prof. dr. Shalom Hakkert and prof. dr. Tom Van Puyenbroeck, for their careful reviews, constructive comments and suggestions that have improved both the content and style of this thesis.

In addition to the professional contribution of these persons, I would like to extend my appreciation to the whole IMOB staff for creating an excellent study environment and kindly help during these four years. Special thanks go to

Nadine Smeyers for helping me with the Dutch summary of this thesis, and Kristel Hertogs who kindly helped me with all administrative matters.

Evidently, my deepest gratitude and love go to my parents, elder sister, and in-laws for supporting me at all times. Last, but certainly not least, I would like to thank my beloved wife Qiong Bao, who has been a source of strength, courage and inspiration. Thanks for always being there.

Yongjun Shen

May 10th, 2012

Summary

Road traffic injuries and fatalities have nowadays been recognized as one of the most important public health issues that requires concerted efforts for effective and sustainable prevention. Given the fact that more and more countries are taking steps to improve their road safety situation, there is a growing need for a country to evaluate its own road safety performance, to compare it with that of other countries, and moreover, to learn from those best-performers as a basis for developing their own road safety policy. In this dissertation research, we implemented road safety product benchmarking and road safety programme benchmarking based on road safety risk indicators and safety performance indicators (SPIs), respectively for 28 European countries. The technique of data envelopment analysis (DEA), originally developed to assess the so-called relative efficiency of a homogeneous set of decision making units on the basis of multiple inputs and multiple outputs, was investigated and applied throughout this dissertation. Various extensions to the methodology were explored to answer the specific research questions that were associated with both road safety benchmarking studies. Useful insights were gained from benchmarking analyses, and valuable recommendations were given to road safety policymakers by indicating practical targets and formulating action priorities to enhance the level of road safety.

In the road safety product benchmarking, we investigated different road safety final outcomes (such as road fatalities). The corresponding road safety risk indicators based on different measures of exposure as well as their evolution over time were compared between countries. Specifically, we developed a DEA-based road safety model (DEA-RS) to evaluate the overall road safety risk of the 28 European countries by simultaneously considering three main risk indicators (i.e., the number of fatalities per million inhabitants, the number of fatalities per 10 billion passenger-kilometres travelled, and the number of fatalities per million passenger cars). That way, the 'efficiency' of each country's current operations was identified. Moreover, by performing clustering analysis to group countries with inherent similarity in their practices, we further applied a categorical DEA-RS model to identify best-performing and underperforming countries in each

group. Useful benchmarks were then identified and a set of practical targets in terms of road fatalities assigned for those underperforming countries.

Furthermore, to capture the dynamic road safety development in each country, we applied the Malmquist productivity index to assess the road safety performance change of countries over time, in which we not only focused on the evolution of road safety final outcomes within a given period, but also took the changes in exposure in the same period into account. It therefore provided more objective results than the ones based on the traditional indicator that only measures the percentage change in road fatalities. Moreover, the decomposition of the index into efficiency change (or catch-up effect) and technical change (or frontier-shift effect) further provided valuable information on whether the improvement in road safety of each country was attained through country-specific progress relative to the other countries that were considered, or just through an overall improvement in the technological environment.

In addition, we also investigated the possibility to take a larger picture of the impact of road crashes into account by including the number of serious injuries as an additional indicator of road safety final outcome to perform road safety product benchmarking and further analyzed its impact on the countries' ranking. In doing so, different types of weight restrictions were formulated in the DEA-RS model to indicate the relationship between road fatalities and serious injuries. Interesting results were obtained inspiring us to apply this kind of model to a more complete road safety product benchmarking practice in the future.

With respect to the road safety programme benchmarking, which is to compare the human-vehicle-infrastructure performance between countries with the purpose of explaining more detailed aspects of crash causation and injury prevention, safety performance indicators situated on the level of intermediate outcomes of road safety were studied, and the combination of individual indicators into a composite road safety performance index was the main focus of this research. Specifically, based on the identification of six leading road safety risk factors (i.e., alcohol, speed, protective systems, vehicle, road, and emergency medical services) within the three main road transport components (i.e., road user, vehicle and infrastructure), we developed a comprehensive set of hierarchically structured SPIs for capturing the road safety performance of a country, and various international data sources providing indicator values for a large set of countries were consulted. Totally, 32 quantitative SPIs were specified with available data collected (or calculated) for 28 European countries, and necessary data processing procedures (including outlier detection and missing data imputation) were performed.

Moreover, to measure the multi-dimensional concept of road safety performance which cannot be captured by a single indicator, we investigated the use of DEA to construct a composite road safety performance index for cross-country comparison. In doing so, a multiple layer DEA-based composite index model (MLDEA-CI) was proposed for hierarchical structure assessment. Based on this model, the most optimal road safety performance index score for each of the 28 European countries was determined by combining all the 32 hierarchical SPIs. Best-performing countries were distinguished from underperforming ones and countries were ranked subsequently. A clear link with the overall road safety risk from the view of the final outcome level was verified. Moreover, country-specific benchmarks were identified for the underperforming countries, and useful insight in the areas of underperformance in each country was gained by analyzing the indicator weights allocated in each layer of the hierarchy. The results enabled policymakers to prioritize their actions to improve the level of road safety in their country.

In addition, for the sake of meaningful and reliable benchmarking, two practical challenges related to data (including missing values and qualitative indicators) were explored in the development of a composite road safety performance index. Regarding the influence of the existence of missing data in the data set on the final index score of 28 European countries, we replaced them by approximations in the form of intervals deduced from multiple imputation in which the true values are believed to lie. An interval MLDEA-based CI model was subsequently applied to obtain for each country an upper and a lower bound of its index score corresponding to its most favorable and unfavorable option, respectively. The interval instead of the precise index score for each country highlighted the underlying imperfect nature of the indicator data, and provided us with a more credible representation of a country's overall road safety performance. Furthermore, we investigated two approaches within the DEA framework for modeling qualitative (or ordinal) data in the context of composite index construction. They are the imprecise DEA-based CI model and the fuzzy DEA-based CI model. A crisp index score was achieved for each country by using the IDEA-based CI model, which is easy for interpretation and use, while in the FDEA-based CI model, fuzzy index scores were obtained based on different possibility levels, which are powerful on the other hand in capturing the uncertainties associated with human thinking. The high similarity of the ranking result based on these two models verified its robustness and implied the reliability of using either of these two approaches for modeling qualitative data.

To conclude, inter-national benchmarking of road safety performance and development is a promising step to improve a country's road safety level. We identified in this dissertation the main research issues with respect to road safety product and programme benchmarking based on different types of road safety indicators, and developed corresponding approaches to deal with these issues. This research mainly contributed to the literature on using the technique of DEA and its various extensions to implement meaningful road safety benchmarking practices. Although it is mathematical in nature, the theory behind it is straightforward and it is currently ready for implementation at the practical level. In addition, from the road safety policy point of view, based on the recommendations with respect to both target setting and action prioritizing from the benchmarking studies described in this dissertation, learning about best practices applied in country-specific benchmarks and (re)formulating concrete road safety strategies and programmes constitute the first next step for each country to take, which in turn, generates new challenges and opportunities for future research.

List of Abbreviations

AHP	Analytic Hierarchy Process
BAC	Blood Alcohol Concentration
BCC	Banker-Charnes-Cooper
BOD	Benefit Of the Doubt
CCR	Charnes–Cooper–Rhodes
CEM	Cross-Efficiency Matrix
CI	Composite Index
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DEA-CI	DEA-based Composite Index model
DEA-MI	DEA-based Malmquist productivity Index
DEA-RS	DEA-based Road Safety model
DMUs	Decision Making Units
EC	European Commission
EFFCH	EFFiciency CHange
EMS	Emergency Medical Services
ERF	European Road Federation
ETSC	European Transport Safety Council
EU	European Union
EuroNCAP	European New Car Assessment Programme
EuroRAP	European Road Assessment Programme
FDEA	Fuzzy DEA model
GDP	Gross Domestic Product
HCU	Hypothetical Composite Unit
IDEA	Imprecise DEA model
IRTAD	International Road Traffic and Accidents Database
MI	Multiple Imputation
MLDEA	Multiple Layer DEA model
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
OR/MS	Operations Research/Management Science
pkm	passenger-kilometres travelled

RQ	Research Question
SBM	Slacks-Based Measure of efficiency
SBR	Seat Belt Reminder
SFA	Stochastic Frontier Approach
SI	Serious Injuries
SPIs	Safety Performance Indicators
s.t.	subject to
TECHCH	TECHnical CHange
UN	United Nations
VRS	Variable Returns to Scale
WHO	World Health Organization

Table of Contents

List of Tables

List of Figures

Chapter 1 General Introduction

This introductory chapter provides a brief background of this dissertation research and defines the research questions as well as the methodology used in this research. The outline of the dissertation is presented at the end.

1.1 Background

The transport sector is an important component of today's world economy directly impacting on the development of our present society and the welfare of human beings. As one of the most fast growing sectors in the post-crisis socioeconomic context, transport systems are expected to experience an accelerated expansion in the next decades due to ever increasing population, rapid motorization, and rising incomes. Projections indicate that by the year 2050, there will be around 3 to 4 times as much global passenger mobility as at the beginning of this new millennium and 2.5 to 3.5 as much freight activity [Organization for Economic Co-operation and Development/International Transport Forum, 2011a]. However, rapid growth of traffic volume, especially motorized road mobility, has also resulted in continuously increasing safety problems. Road safety is important not only because of the lost travel time or cost of property damage, but mainly because of the loss of human life and serious injuries sustained. Since the first death involving a motor vehicle which is said to have taken place in London in 1896, road traffic crashes have claimed an estimated 40 million lives up to now, and many more suffer non-fatal injuries [World Road Association, 2003]. These not only lead up to reduced worker productivity and trauma affecting a victim's private life, but also cause great emotional and financial stress to the millions of families affected. More seriously, in most regions of the world, especially for those low- and middle-income countries, this hidden epidemic is still spreading (see Table 1.1).

As for the high-income countries, such as those in Europe, which has been recognized as one of the safest road traffic regions in the world, they also suffer from the road crash problem. Due to the high level of car ownership, road transport has emerged as the dominant segment in Europe's transport sector accounting for roughly 84% of all passenger transport and 47% of freight transport [European Commission, 2011a]. However, it is also responsible for the majority of negative impacts on safety, which accounts for over 100 times more deaths than all other transportation modes (rail, air, maritime, etc.) together [Forum of European Road Safety Research Institutes & European Conference of Transport Research Institutes, 2009].

Table 1.1 Predicted road traffic fatalities by region (in thousands), 1990-2020

Region*	No. of countries	1990	2000	2010	2020	Change (%) 2000-2020	Mortality rate (Fatalities/million inhabitants)	
							2000	2020
East Asia & Pacific	15	112	188	278	337	79	109	168
Eastern Europe & Central Asia	9	30	32	36	38	19	190	212
Latin America & Caribbean	31	90	122	154	180	48	261	310
Middle East & North Africa	13	41	56	73	94	68	192	223
South Asia	7	87	135	212	330	144	102	189
Sub-Saharan Africa	46	59	80	109	144	80	123	149
Sub-total	121	419	613	862	1124	83	133	190
High-income countries	35	123	110	95	80	-27	118	78
Total	156	542	723	957	1204	67	130	174

* Data are displayed according to the regional classifications of the World Bank.

Source: World Health Organization (2004)

In 2010, nearly 31,000 people died in the 27 Member States of the European Union (EU-27) as a consequence of road traffic crashes. Around 300,000 were seriously injured and many more suffered slight injuries [European Transport Safety Council, 2011]. Despite the fact that the number of road fatalities keeps decreasing over the last several decades, it is, however, still far away from the 27,000 objective for 2010 [European Commission, 2001; 2003] (see Figure 1.1). Involvement in road crashes remains as one of the leading causes of death and hospital admission for EU citizens under 45 years of age [European Commission, 2009]. Moreover, the huge costs in health services and the added burden on public finances due to road traffic injuries and fatalities representing approximately 130 billion Euro, or over 1% of the EU Gross Domestic Product

(GDP) in 2009 [European Commission, 2010a], have also become increasingly socially unacceptable and difficult to justify to citizens. Road safety therefore continues to be a priority area for action of the EU.

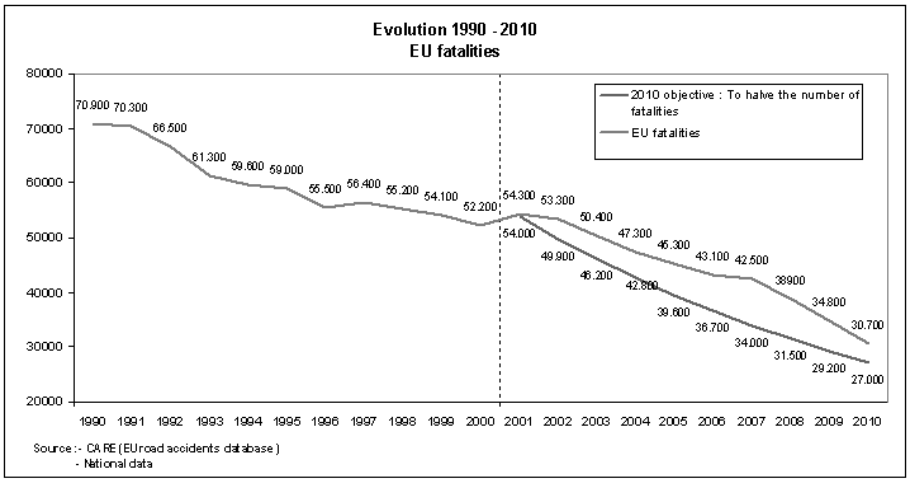


Figure 1.1 Evolution of road fatalities in the EU-27 (1990-2010)

Source: European Commission (2011b)

On the other hand, the road traffic crashes and consequent injuries and fatalities, traditionally regarded as random, unavoidable ‘accidents’, have been more and more recognized as a preventable public health problem due to a better understanding of the nature of crashes over the past decades. As a result of this shift in perception, road traffic crashes and their health implications have demanded the attention of decision-makers all over the world and safety policy has been firmly placed in the public health arena. Under these circumstances, a large number of road safety strategies and programmes have been launched and put into effect at either a national, regional, or even global level. Worldwide, the United Nations proclaimed the period 2011 to 2020 as the ‘Decade of Action for Road Safety’ in May 2011. International cooperation for making road safety a priority is advocated with the purpose of 50% reduction in road fatalities and injuries on the predicted global death toll by 2020. Analyses conducted by the Global Road Safety Facility indicate that achieving this target would result in the saving of an estimated 5 million lives and 50 million serious injuries requiring hospitalization being avoided, with an estimated saving of more than US $3 trillion (see also Figure 1.2) [Guria, 2009].

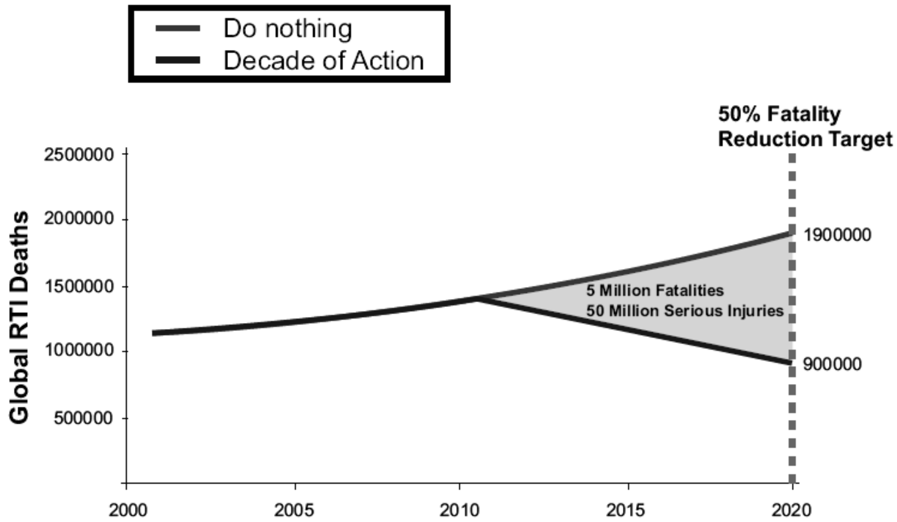


Figure 1.2 Changing direction: Potential of a decade of action for road safety

Source: Guria (2009)

In Europe, at the early 21st century, the EU has already set itself an ambitious target of reducing the number of road fatalities by half during the past decade [European Commission, 2001]. Although the initial target was not met by the end of 2010 (see Figure 1.1), the action has been a strong catalyst of efforts made by Member States to improve their road safety. Furthermore, in the European 'Policy Orientations on Road Safety 2011-2020' [European Commission, 2010a], the Commission has proposed to continue with the target of halving the total number of road fatalities in the EU by 2020, which is apparently more challenging than the previous one yet gives a clear signal of Europe's commitment towards road safety. On the national level, an increasing number of countries begin or continue to implement long term road safety strategies towards their reduction or eventual elimination of road traffic injuries and fatalities, such as the Sustainable Safety concept in the Netherlands [Wegman & Aarts, 2006] and the Swedish Vision Zero [Organization for Economic Co-operation and Development/International Transport Forum, 2008a].

Although more and more countries are taking steps to improve their road safety situation, they work in most cases on their own to tackle their specific road safety problems. This is right to a large extent because the socioeconomic conditions, the motorization levels, and the road safety experiences are different from country to country and from region to region. However, for those countries within the same region or that have already passed through similar stages of challenges and development, such as the EU Member States, there are quite a number of common problems that can be identified in a close cooperation, and improvement can be expected by learning lessons from existing best practices in other countries (even if the final solutions or priorities could be different from one country to another in accordance with their own safety characteristics). Consequently, comparison between a range of countries in terms of their road safety performance and development or – using state-of-the-art terminology – *inter-national benchmarking of road safety*, is currently widely encouraged and advocated by governments, donors, practitioners, planners, and researchers for the purpose of better understanding each country's relative safety situation, and moreover, trying to learn from those better-performing countries in terms of setting practical targets, designing effective strategies, determining intervention priorities, monitoring programme effectiveness, and ultimately, achieving its own safety objectives.

1.2 Road Safety Benchmarking

1.2.1 The concept of benchmarking

The term *benchmarking*, originally derived from the work of cobblers who would place someone's foot on a 'bench' and mark it out to make the pattern for the shoes, was firstly invented in the private sector as a tool for improving various operations by establishing a point of reference by which it is possible to judge quality, value or other important factors. Now, the concept of benchmarking is further extended and widely adopted in both profit and non-profit organizations. One of the operational definitions of benchmarking is:

> "*the process of continuously measuring and comparing ones business processes against comparable processes in leading organizations to obtain information that will help the organization identify and implement improvements.*" [American Productivity and Quality Center, 1993].

First and foremost, benchmarking is a systematic comparison of the process and performance of one production entity against other entities, which could be countries, organizations, firms, industries, divisions, projects, or individuals. Moreover, the essence of benchmarking is the process of identifying the highest standard of excellence for products, services, or processes, and then making the improvements necessary to reach those standards – commonly known as 'best practice' [Bhutta & Huq, 1999]. In addition, benchmarking does not represent the end of the process, but is an ongoing diagnostic management tool focused on learning, collaboration and leadership to achieve continuous improvement in the organization over time [Garlick & Pryor, 2004].

Benchmarking is a versatile tool that can be applied in a variety of ways to meet a range of requirements for improvement. It can firstly be used to make intra-organizational comparisons, which involves benchmarking against internal operations or standards, usually in a multidivision or multinational enterprise. Benchmarking can also be – and most frequently is – used to make inter-organizational comparisons. It deals with benchmarking against other entities in the same context, no matter whether they are direct competitors or not. In addition, benchmarking can also be used to make longitudinal comparisons, where the performance of one or more production entities in different time periods is compared.

Since the first successful application implemented by Xerox Corporation in the late 1970s, benchmarking quickly became one of the fastest growing techniques for quality and performance improvement and has been receiving significant attention in a multitude of entities engaged in a variety of performance evaluation, quality management, and continuous improvement activities [Camp, 1989; Spendolini, 1992; Andersen & Pettersen, 1996; Elmuti et al., 1997; Keehley et al., 1997; Srinivas, 2000; Garlick & Pryor, 2004; Geraedts & Selbmann, 2004; Lau et al., 2005; Luu et al., 2008; Cheng et al., 2009; Chung, 2011; Lai et al., 2011].

In terms of road safety, more and more countries have recognized the importance of benchmarking practices in improving their level of road safety, especially the inter-national comparisons. Taking the EU as an example, the European Commission has claimed that "the establishment of a structured and coherent cooperation framework which draws on best practices across the Member States, [i]s a necessary condition to implement in an effective manner the road safety policy orientations 2011-2020." [European Commission 2010a]. An instructional definition on inter-national benchmarking of road safety is given as:

> "*a process in which countries evaluate various aspects of their performance in relation to other practices, among which the so-called 'best in class'. The benchmark results enable countries to learn from others as a basis for developing measures and programmes which are aimed at increasing their own performance.*"
> [Wegman and Oppe, 2010].

1.2.2 Benchmarking process

To implement benchmarking, a number of different process models have been proposed during the past decades describing the steps of a benchmarking study. One such model is the benchmarking wheel [Andersen, 1995], as shown in Figure 1.3.

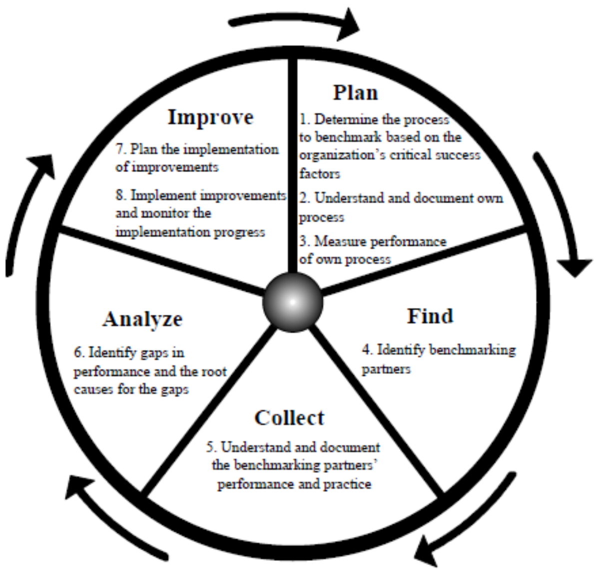


Figure 1.3 The benchmarking wheel

Source: Anderson (1995)

The main content of each of the five phases in a typical benchmarking study is:

Phase 1 Plan. Prepare the benchmarking study by laying the groundwork for the coming phases, such as selecting the process to be benchmarked and thoroughly understanding how that process is performed within one's own organization.

Phase 2 Find. Identify benchmarking partners and obtain acceptance for their participation in the study.

Phase 3 Collect. Perform the same thorough documentation of the benchmarking partners' process as was done for one's own in the plan phase.

Phase 4 Analyze. Find gaps between the performance of one's own process and that of the benchmarking partners, and also determine the root causes for these gaps in practice.

Phase 5 Improve. Implement improvements based on the findings from the observation and analysis of the benchmarking partners. The outcomes can be used for the next benchmarking study with the purpose of continuous improvement.

Figure 1.4 The road safety benchmarking cycle

Adapted from Wegman et al. (2008)

In the road safety context, a similar benchmarking cycle can be considered consisting of the following core activities (see Figure 1.4): determining the key components for road safety benchmarking, identifying the benchmarking partners (or countries), constructing indicators for meaningful comparisons and data gathering, examining gaps in performance and their root causes, and finally, establishing future attainable performance and monitoring progress. Each of these five activities poses different challenges for the benchmarking organization, and all of them are vital elements in a complete road safety benchmarking study. In this dissertation research, most of these five activities will be investigated except for the intervention and monitoring in the final phase of the cycle.

1.2.2.1. Determining the key components for road safety benchmarking

To compare the road safety performance between countries, we should always determine what to benchmark in the first place. In this respect, Eksler (2009) proposed a so-called process and performance benchmarking framework for road safety management, which is presented in Figure 1.5.

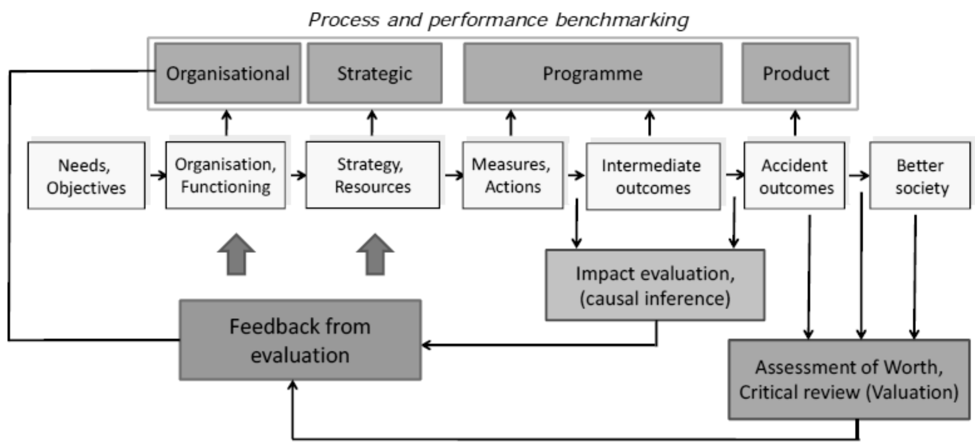


Figure 1.5 Process and performance benchmarking in road safety management
Source: Eksler (2009)

In this comprehensive benchmarking framework, four aspects of the road safety management and improvement process have been identified. They are: organization, strategy, programme and product. More specifically, *product benchmarking* is used to compare road safety final outcomes, such as road traffic mortalities. So far, most of the road safety benchmarking studies have focused on this aspect. *Programme benchmarking*, which is used to compare activities related to human-vehicle-infrastructure performance, such as drink driving, seat belt wearing, vehicle and road safety ratings, and corresponding policy action, has also been given more attention in current road safety studies since they are causally related to crashes or injuries and can provide a better understanding of the process that leads to crashes. Worldwide, the two most representative benchmarking studies concerning the above two aspects are the 'IRTAD Road Safety Annual Report' and the 'Country Reports on Road Safety Performance' conducted within the OECD vision [International Traffic Safety Data and Analysis Group, 2012; Organization for Economic Co-operation and Development/ International Transport Forum, 2008b].

The remaining two aspects, i.e., *strategic* and *organizational benchmarking*, are used to compare national road safety strategies, resources, management and the organizational framework. However, due to the lack of appropriate indicators characterizing their features, only some initial attempts have been carried out at this moment, such as Al-Haji (2007), Wegman et al. (2008), and Eksler (2009).

In addition, Al-Haji (2007) and Wegman et al. (2008) also proposed the use of a road safety index, which combines performance indicators/indexes developed in the above separate benchmarking aspects into one overall index, and it is named as *integrated benchmarking*. The application of this concept will be further discussed in Section 1.3.

1.2.2.2. Identifying the benchmarking partners

Having determined the subject of the exercise, no matter if it is for road safety product benchmarking, programme benchmarking, or even for strategic and organizational benchmarking, the next step is to identify the benchmarking partners, i.e., with whom to compare. It is not an easy task to define an uniform criterion on the selection of benchmarking partners (or countries) for international road safety benchmarking practices. In a general sense, all the countries are comparable in terms of their road safety performance. However, in order to achieve adequate and meaningful results during comparisons, road safety benchmarking studies usually have to be carried out between similar countries or regions at as much as possible the same level of development, motorization and with a similar type of transport system [Al-Haji, 2007]. For instance, in Europe, the *SUNFlower* study [Koornstra et al., 2002] focused on the three best-performing countries in road safety (Sweden, United Kingdom and the Netherlands) and in the following *SUNFlower+6* study [Wegman et al., 2005], three Southern European countries (Greece, Portugal and Spain, with a special position for Catalonia) and three Central European countries (Hungary, Slovenia and the Czech Republic) were included; The *SECBelt* study [European Transport Safety Council, 2005] worked on road safety causes and problems in the Southern, Eastern and Central European countries; Another ETSC study concentrated on the performances of Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) in different areas of road safety [Eksler et al., 2009]. Moreover, some large-scale benchmarking studies were also carried out within the whole EU vision, such as the *SafetyNet* study [Thomas et al., 2009], the *SUNFlowerNext* study [Wegman et al., 2008], and the ongoing *DaCoTA* study (http://www.dacota-project.eu/).

1.2.2.3. Constructing indicators for meaningful comparisons

The third step for implementing inter-national benchmarking of road safety is to develop a set of relevant indicators for the selected benchmarking component. They can be measured in some common terms such as a rate (e.g., number of fatalities per population), a percentage (e.g., percentage of seat belt usage), or as qualitative information (e.g., level of national road safety intervention: 'low', 'relatively low', 'high', and 'extremely high'). Moreover, indicator values have to be collected for all the countries involved in the benchmarking study. In general, developing appropriate road safety indicators for a specific benchmarking study and structuring them in a logical way is the basis of a successful benchmarking practice. This part will be further elaborated in Section 1.3.

1.2.2.4. Examining gaps in performance and their root causes

In this step, the process knowledge from the previous steps is put together to identify the gaps in road safety performance between the countries under study and to understand the root causes for these gaps. This is the most important step in the entire benchmarking study, but also the most challenging task to fulfill. Today, various benchmarking tools have been developed which range from relatively simple (e.g., using statistical tables and graphs) to highly complex (e.g., index-based approaches, see also Section 1.3) depending on the number of indicators involved, the details of data, and the complexity of techniques used in calculation and analysis. According to Camp (1995), an intuitive way for gap analysis is to present data in some graphical form. These graphics are easy to understand and are capable of illustrating multiple dimensions simultaneously. However, it is a difficult task for the analysts to integrate all the elements into complete and meaningful information. Ratio analysis [Schefcyzk, 1993] is another approach that is commonly used due to its simplicity, such as using the number of fatalities per population to rank countries in a road safety product benchmarking study. One problem with ratios is that there can be several of them (e.g., the number of fatalities per vehicles and the number of fatalities per distance travelled). Comparisons of a single ratio might thereby lead to misclassifications and incorrect judgments. In the applications of multiple ratios, the weighting of the ratios would require the formulation of complex decision rules and their justification, as well as a much greater computational workload. An available solution is to perform multi-criteria decision analysis, and one of the methods belonging to this category is known as analytic hierarchy process (AHP) [Saaty, 1980], which utilizes a weighted scoring method in the analysis of various indicators. It provides a single score using perceptual values as set forth by decision makers. Despite being effective, the main drawback of this method is the involvement of a high degree of subjectivity. The ordinary least squares (OLS) statistical techniques, such as multiple regression, are also widely applied to assess comparative performance of different entities [Hayashi, 2000]. Even though there is a strong theoretical foundation for such statistical tools, their primary limitation is in the underlying assumptions of normality, homoscedasticity, and serial independence of regression residuals. Also, Bessent et al. (1982) indicate that major difficulties arise when the OLS is used in multiple output cases due to the implicit impact on outputs having the same input resources. In addition, it measures a correlation or central tendency rather than best practice. Frontier analysis is one other technique recently receiving significant attention in benchmarking studies. The data envelopment analysis (DEA) and the stochastic frontier approach (SFA) are the two representatives within this field [Coelli et al., 2005; Bogetoft & Otto, 2011]. The SFA uses statistical techniques to estimate a transformation frontier and to estimate efficiency relative to the estimated stochastic frontier [Aigner et al., 1977]. A valuable characteristic of this approach is the introduction of a disturbance term representing noise, measurement error and exogenous shocks beyond the control of the production unit. This phenomenon permits decomposition of the deviation of an observation from the deterministic kernel of the frontier into two components: inefficiency and noise. On the other hand, the method imposes an a priori assumption on the production technology by choosing a functional form (e.g., Cobb-Douglas, translog, etc.), which is risky because most of the distributional characteristics of the production technology are a priori unknown. Moreover, the precise specification of the error structure is difficult, sometimes even impossible to ascertain. Such specification is in fact likely to introduce another potential source of error. Compared with the stochastic parametric frontier approach, the DEA is a non-parametric method imposing no assumptions on the specific statistical distribution of the error terms. It applies mathematical programming methodology to measure the relative efficiency of a homogeneous set of decision making units (DMUs) by constructing an efficient production frontier based on best practice(s) [Charnes et al., 1978]. In doing so, the data are believed to be able to 'speak for themselves' and the specification error is minimized. However, the DEA model does not allow for measurement error or random shocks. Instead, all these factors are attributed to calculate (in)efficiency. In this study, the DEA approach will be adopted as a benchmarking tool to examine the gaps in road safety performance between countries as well as the root causes for those gaps. A more specific description of this technique will be given in Section 1.5.

1.2.2.5. Establishing future attainable performance and monitoring progress

After finishing the analysis, this step performs target-setting for those underperforming countries in terms of different road safety aspects, and also determines what needs to be done to match the best practice and to fill the gaps for the process. Moreover, as a cycle, such a benchmarking practice should be carried out at regular intervals so as to evaluate the results of interventions and to monitor progress on road safety in each country in order to achieve continuous improvement over time.

1.3 Road Safety Indicators and Indexes

As indicated in the previous section, to be able to implement inter-national benchmarking in the field of road safety, a set of indicators that summarizes the country's road safety performance from different benchmarking aspects has to be developed, which serves as the basis for a successful benchmarking process. However, when a number of indicators are considered for a particular benchmarking aspect, simple comparisons per indicator with the purpose of examining the gaps in performance may only show a small piece of the road safety picture, and can be misleading since different countries may operate in different circumstances with different focal points. Consequently, a composite road safety indicator (or index), which combines individual indicator values into one single score, is often computed for the sake of meaningful benchmarking.

1.3.1 Road safety indicators

Traditionally, crash data such as the number of fatalities gathered as part of the routine police procedures are viewed as road safety final outcomes and mostly investigated in road safety studies. Such numbers give an idea about the absolute size of the road safety problem in a country. However, they are not directly comparable between countries. Therefore, the concept of risk, which is defined as the ratio of road safety final outcomes and some measure of exposure (e.g., the population size, the number of registered vehicles, or distance travelled), is often used in the context of benchmarking. However, these final outcomes and the corresponding risk indicators are usually considered as the 'worst case scenario' in the insecure operational conditions of road traffic, and are insufficient in explaining more detailed aspects of crash causation and injury prevention [Vis, 2005]. Today, having recognized the complex character of the road safety phenomenon, countries all over the world are encouraged to develop data collection procedures in terms of road safety indicators to cover not only the final outcomes and the exposure measures, but also intermediate outcomes (e.g., levels of mean traffic speeds, seat belt wearing, drink driving, and vehicle and infrastructure safety ratings), institutional delivery outputs (including different categories of enforcement effort), and socio-economic costs associated with road trauma as well [Organization for Economic Co-operation and Development/International Transport Forum, 2008a]. Over the last decade, a large number of road safety indicators have been developed and increasingly used as a supportive instrument for inter-national (or inter-regional) comparisons and monitoring of road safety progress (e.g., European Transport Safety Council, 2001; Vis, 2005; Al-Haji, 2007; International Organization for Standardization, 2008; Wegman et al., 2008; Hermans, 2009a; Gitelman et al., 2010).

Particularly, a road safety target hierarchy was proposed for the development of various indicators. The concept originated in New Zealand [Land Transport Safety Authority, 2000], and further used in the European *SUNflower* study [Koornstra et al., 2002] and the European *SafetyNet* study [Thomas et al., 2009] as well. Now, it has also become the theoretical basis for the creation of the European Road Safety Observatory with the purpose of bringing together all Community activities in relation to safety data and knowledge. In general, the target hierarchy describes road safety as a pyramid consisting of five vertical layers as presented in Figure 1.6.

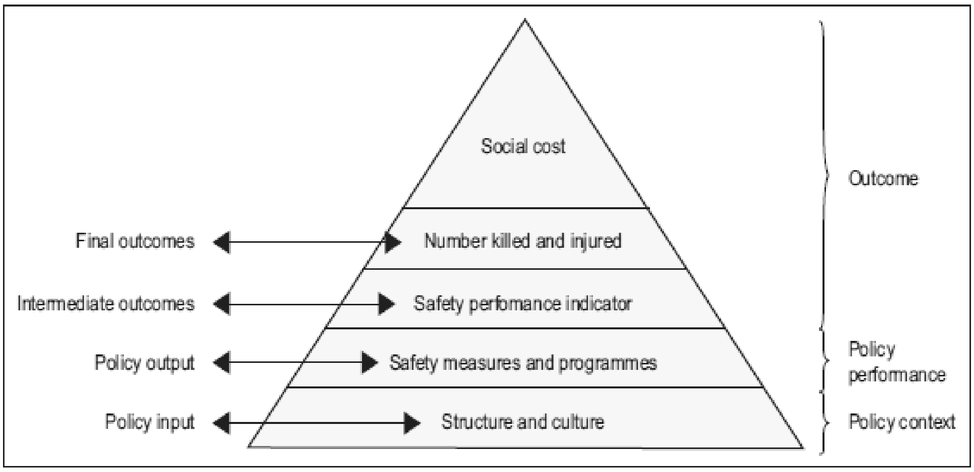


Figure 1.6 Target hierarchy for road safety

Source: Wegman et al. (2005)

From bottom to top they are: structure and culture (describing the background conditions of a country or its policy context); safety measures and programmes (or the road safety policy performance); safety performance indicators (also known as intermediate outcomes); the number of fatalities and casualties (as the final outcomes); and the social costs due to crashes and injuries at the very top.

The layers of the pyramid are stacked simply but logically. They imply the causal relationship between indicators at the different layers. For instance, policy interventions such as a high frequency of road side alcohol check, will first have to result in a decreasing rate on drinking and driving before it can be made credible that the intervention has an effect on reducing alcohol-related crashes and risks. In other words, the pyramid enables us to better understand the development at the top by explaining the change at the bottom.

Apart from this vertical dimension, two other dimensions should also be considered based on this pyramid (not depicted in Figure 1.6) [Morsink et al., 2007]. At the horizontal level, road safety problems can be specified in a disaggregated way such as per road user group, transport mode, or road type. Comparisons can then be conducted in a country or between countries. The third dimension is time allowing to show the development of factors in both the horizontal and vertical dimension over time. Based on this 3-dimensional framework, relevant road safety indicators can be formulated at each layer as a basis for implementing inter-national benchmarking of road safety performance and development.

1.3.2 Road safety indexes

Different from separate indicators, a composite indicator or index (CI) is a mathematical aggregation of a set of individual indicators that measures multi-dimensional concepts but usually has no common units of measurement [Saisana & Tarantola, 2002]. According to Saisana & Tarantola (2002) and the Organization for Economic Co-operation and Development (2008), the main pros and cons of using CIs are summarized as in Table 1.2.

Table 1.2 Pros and Cons of Composite indexes

Pros	Cons
– Enable users to compare complex and multi-dimensional realities effectively and summarize them in view of supporting decision-makers – Are easier to interpret than trying to find a common trend in many separate indicators – Can assess progress of countries over time – Reduce the visible size of a set of indicators without dropping the underlying information base – Make it possible to include more information within the existing size limit – Place issues of country performance and progress at the centre of the policy arena – Facilitate communication with general public (i.e. citizens, media, etc.) and promote accountability – Help to construct/underpin narratives for lay and literate audiences	– May send misleading policy messages if poorly constructed or misinterpreted – May invite simplistic policy conclusions – May be misused, e.g. to support a desired policy, if the construction process is not transparent and lacks sound statistical or conceptual principles – The selection of indicators and weights could be the subject of political dispute – May disguise serious failings in some dimensions and increase the difficulty of identifying proper remedial action – May lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored

In general, "… it is hard to imagine that debate on the use of composite indicators will ever be settled…" [Saisana et al., 2005]. However, in case the methodological aggregation process is sound and the results clear, the construction of a CI over a set of indicators is worthwhile, and it can be utilized as a powerful benchmarking tool in policy analysis and public communication. During the last decade, a large number of CIs have been developed by various national and international organizations including United Nations (UN), Organization for Economic Cooperation and Development (OECD), World Health Organization (WHO), World Bank, and European Commission (EC), amongst others, involved in wide ranging fields such as economy, society, governance, security, environment, sustainable development, globalization and innovation [Saisana & Tarantola, 2002; Freudenberg, 2003; Munda, 2005; Organization for Economic Co-operation and Development, 2008; Singh et al., 2009]. According to a comprehensive review by Bandura (2008), around 180 different CIs have been identified all over the world. Some of them are listed as follows:

- Human Development Index – UN;
- Programme for International Student Assessment – OECD;
- Overall Health System Achievement Index – WHO;
- Governance Indicators – World bank;
- Innovation Capacity Index – World Economic Forum;
- Global Competitiveness Index – World Economic Forum;
- Globalization Index – World Markets Research Centre;
- Country Risk Rating – World Markets Research Centre;
- Internal Market Scoreboard and Internal Market Index – EC;
- Science and Technology Indicators – EC;
- Environmental Sustainability Index – Columbia University and Yale University.

Compared to other research fields, the development of a composite index for road safety benchmarking is relatively new, since the traditional studies mainly focus on the road safety final outcomes, and ratio analysis is commonly conducted, such as using the number of fatalities per head of population to assess the relative road safety situation of a country. Nowadays, since more and more indicators are developed describing the complex character of the road safety phenomenon, simple ratio analysis no longer satisfies the need of modern road safety benchmarking practices. Recently, several studies were carried out aiming at the development of a composite road safety index which enabled meaningful national or sub-national comparison and monitoring of road safety performance.

Specifically, Al-Haji (2007) suggested a road safety development index (RSDI) which consists of three focuses of the road safety domain. They were product focus (fatality rates), people focus (road user behaviour), and system focus (safer vehicles, safer roads, socio-economic level, enforcement, and organizational performance). The index was then applied for the comparison of road safety progress in both highly motorized countries (eight European countries) and less motorized countries (five Southeast Asian countries). Although different numbers of road safety indicators were selected representing the above three focuses in each empirical study due to data availability, one composite index was expected for both sub-studies. For this purpose, four weighting methods were adopted, which were equal weighting, expert judgments, subjective weights based on previous experience, and principal component analysis. The empirical and theoretical assessments indicated that the proposed RSDI could give a broader picture of the road safety situation in a country and could serve as a simple and easily understandable tool for policy makers and the public.

In the *SUNflowerNext* study [Wegman et al., 2008], three different types of performance indicators were distinguished, which were road safety performance indicators (i.e., the top three layers of the pyramid in Figure 1.6), implementation performance indicators (dealing with different components of causal relationships between the different layers of the pyramid, such as between the changes in safety performance indicators and changes in the number of casualties), and policy performance indicators (i.e., the second layer of the pyramid from the bottom). Moreover, a composite road safety index combining the indicators in each layer of the pyramid was explored. Two weighting schemes, i.e., principal component analysis and factor analysis, were examined based on the data collected for 27 European countries. The analysis revealed that such an index gave a more enriched picture of road safety and the countries' ranking based on the combination of different indicators was not necessarily similar to the traditional ranking of countries based only on mortality or fatality rates.

Hermans (2009a) explored a methodological framework for developing a composite road safety performance index for cross-country comparison. The following steps were distinguished: selecting indicators, collecting indicator data, univariate analysis, multivariate analysis, weighting, aggregation, robustness testing, and computing, evaluating and visualizing final index scores. This study could be considered as a valuable guideline for future research on developing composite road safety indexes. Moreover, to illustrate the use of this framework, six risk factors, i.e., alcohol and drugs, speed, protective systems, vehicle, roads, and trauma management, were considered, one safety performance indicator for each risk factor was defined, and five weighting approaches were investigated to combine the separate indicators into one overall index for 21 European countries, which were: factor analysis, budget allocation, analytic hierarchy process, data envelopment analysis, and equal weighting [Hermans et al., 2008]. The results were further compared with one of the road safety risk indicators, which was the number of fatalities per million inhabitants. The study concluded that comparing the performance of countries in terms of road safety by means of an index at the intermediate outcome level enabled earlier and goal-oriented action.

All the studies mentioned above clearly demonstrate the necessity and feasibility of creating a composite road safety index for road safety benchmarking purposes among a set of countries. However, research attention still needs to be paid to some theoretical and practical aspects of the road safety index construction. First of all, Al-Haji (2007) and Wegman et al. (2008) implemented integrated benchmarking by combining indicators in all the different benchmarking aspects (i.e., product benchmarking, programme benchmarking, strategic benchmarking, and also organizational benchmarking) in one road safety index. The idea itself is attractive, however, it is well accepted and proved that underlying causal relationships exist between different benchmarking components. Integrating all these components thereby implies to combine indicators that are actually inter-dependent, which is not in accordance with the basic principle of index construction and will generate a problem of double or even triple counting the effect of one factor in the final index. This obstacle restricts to a great extent the application of this concept unless a new theoretical framework appears.

In Hermans (2009a), only the indicators belonging to the intermediate outcome layer of the pyramid were considered for the purpose of programme benchmarking. The created index score, however, was merely compared with one road safety risk indicator, which is insufficient in revealing their causal relationship. In other words, additional risk indicators (using different measures of exposure and other final outcomes) should also be taken into account in the product benchmarking area. Moreover, in Hermans (2009a), a relatively small number of road safety performance indicators were considered (i.e., one quantitative indicator for each risk factor), which is also insufficient to completely measure the entire situation of all the risk factors. In the other two studies, although one or several indicators were suggested for each aspect constituting a hierarchical structure, all the indicators were actually treated as in the same layer under most of the weighting schemes. In fact, the structure of the indicators contains valuable information worthwhile to be considered in index construction.

With respect to the weighting schemes adopted in all three studies, no matter if they are objective or subjective ones, most of them (except for DEA) assign the same indicator weights for all the countries under study. It indeed enables the comparison among countries on a common base. However, in that way, we make no use of country-specific characteristics. In other words, the importance level of each indicator in each country is ignored, which makes the examination of root causes of poor performance in each country difficult. In this respect, the DEA approach, which is based on self appraisal, has much to be recommended. By using this technique, each country obtains its own best possible indicator weights. Thus, key problems on road safety can be identified for each country separately, and policy-makers could not complain about unfair weighting, because the highest possible index value is obtained and any other weighting scheme would generate lower composite scores.

Furthermore, due to data unavailability, all these three studies only assessed the road safety performance of countries in one year. However, research on time series data collected at regular intervals would help in gaining a clear understanding of trends and expected progress towards postulated targets and benchmarks' performance. Apart from this, there are still some issues related to data which need to be carefully handled in the development of road safety indexes, two of which are missing values and qualitative indicators. Specifically, obtainment of measurable and quantitative indicators is commonly the prerequisite of any index research. However, this becomes more and more difficult since the natural uncertainty of reality often leads up to imprecision and vagueness inherent in the information that can only be represented by means of qualitative data, such as the policy performance indicators used in Wegman et al. (2008). Simply treating them as quantitative ones could result in wrong conclusions. Moreover, an extension of the data set used for road safety index research raises the issue of missing values, which to a great extent restricts researchers from performing classical analyses as complete data matrices are usually required. Consequently, how to effectively settle these data problems directly affects the result of road safety index research and the success of the benchmarking as well.

Finally, it is necessary to mention here that benchmarking is mainly used as a tool for learning from each other. Therefore, obtainment of final index scores should not be the only interest, but the background of those scores, the factors that contribute to the scores, and the potential for improvement.

1.4 Objective and Research Questions

The main objective of this dissertation research is to perform comprehensive inter-national benchmarking on the road safety product and programme, respectively, so that countries can better understand their own relative road safety situation, and moreover, can learn from those better-performing countries as a basis for developing their own road safety policy. For this purpose, different road safety indicators have to be specified for each component, and indexes are then constructed for meaningful benchmarking. In doing so, the research challenges mentioned in the previous section need to be carefully taken into account, and a set of research questions can then be formulated as follows, which is actually the guiding link throughout this dissertation.

In the road safety product benchmarking study, the main research questions are:

RQ1: How to obtain an overall picture of a country's road safety performance when different final outcome risk indicators are considered?

RQ2: How to set practical targets for similar underperforming countries in terms of the number of road fatalities?

RQ3: How to assess road safety performance change over time by taking both final outcome evolution and exposure change into account?

RQ4: What is the impact of utilizing other final outcomes, e.g., serious injuries, in addition to fatalities for road safety product benchmarking?

In the road safety programme benchmarking study, the main research questions are:

RQ5: Which are current available national safety performance indicators and how can they best be structured?

RQ6: How to reflect a layered hierarchy of indicators in constructing a road safety performance index and what is the added value?

RQ7: How to obtain a reliable index score for each country when missing data exist?

RQ8: What is the possible way to incorporate qualitative indicators?

1.5 Data Envelopment Analysis

To achieve the main research objective, the technique of data envelopment analysis (DEA) – one of the powerful benchmarking tools currently receiving considerable attention in the Operations Research/Management Science (OR/MS) literature – is investigated and applied throughout this dissertation with its various extensions to answer the specific research questions mentioned above. In this section, we briefly introduce the historical development of this technique, its fundamental mechanisms, and different model formulations, based on which the main strengths and limitations of this technique are discussed, and finally, some important methodological extensions for this dissertation research are outlined.

1.5.1 Brief history of the method

The term *data envelopment analysis* was first reported in the *European Journal of Operations Research* by Charnes, Cooper and Rhodes in 1978 based on Rhodes' PhD dissertation research entitled 'A DEA Approach to Evaluation of the Program Follow Through Experiment in U.S. Public School Education'. It was the failure of using all the statistical-econometric approaches that led Rhodes to suggest Farrell (1957)'s work 'The measurement of productive efficiency' as an alternative for analyzing efficiency (E) as a measure of performance expressed in the form of a ratio as follows:

$$E=\frac{\text{Output}}{\text{Input}} \tag{1-1}$$

Farrell identified two components of efficiency: a *technical efficiency*, which showed the ability to maximize output from a given input, and a *price efficiency*, which reflected the use of different inputs allocated in optimal proportions (and hence also referred to as *allocative efficiency*). Considering both measures together thus provided an *overall* (or *economic*) *efficiency*. However, Farrell's empirical work had been confined to single-output cases.

Building on the ideas of Farrell (1957), the seminal work "Measuring the efficiency of decision making units" by Charnes, Cooper & Rhodes (1978) provided a new approach of obtaining empirical estimates of relations between multiple inputs and multiple outputs by constructing an efficient production frontier and assessing the so-called *relative efficiency* for a set of entities, referred to as decision making units (DMUs), which has subsequently been titled data envelopment analysis or DEA.

In microeconomic production theory, a firm's input and output combinations are depicted using a *production function*, also known as *efficient production frontier*. Such a frontier indicates the maximum quantity of outputs that can be obtained from a given combination of inputs. At the same time, it also expresses the minimum quantity of inputs that must be used to achieve a given output level [Seiford & Thrall, 1990; Coelli et al., 2005]. To construct this frontier, a set of observations that expresses the output level obtained by applying a specific combination of input production factors is needed. In the context of DEA, the observations correspond to the homogeneous DMUs being evaluated.

A fundamental assumption behind this method is that if a DMU can produce a certain level of output utilizing specific input levels, other DMUs of equal scale should be capable of doing the same if they were to operate efficiently. Thus, the heart of the DEA analysis lies in finding the 'best' DMU(s) which are viewed as the most efficient under the given circumstances, and are used to construct the efficient production frontier. The others that either make less outputs with the same inputs or make the same outputs with more inputs are inefficient, and the degree of their inefficiency can be measured based on the distance from the frontier. In this way, the best-performer(s) can be set as the 'benchmark' for others to aspire to. Meanwhile, it indicates the potential for improvement under current conditions.

Since the initial study by Charnes, Cooper, and Rhodes, DEA has been quickly recognized as a powerful analytical research technique for modeling operational processes in terms of performance evaluation (e.g., Cherchye et al., 2008), decision making (e.g., Ertay & Ruan, 2005), and benchmarking (e.g., Hermans et al., 2009b). At the same time, various methodological extensions to the original model and variations to it have been successfully developed and extensively investigated (e.g., Charnes et al., 1994; Zhu & Cook, 2007). Although DEA is not always the right tool for a problem (see Section 1.5.4), its empirical orientation and minimization of a priori assumptions have caught great attention of analysts from different research fields, and has resulted in its applications to a host of different types of entities engaged in a wide variety of activities in many contexts [Cooper et al., 2000, 2004; Zhu, 2003; Emrouznejad et al., 2008; Cook & Seiford, 2009]. By 1992, 14 years later, 472 publications on DEA reflected a growing interest in this technique [Charnes et al., 1994]. By 1999, over 800 citations were recorded, and up to the year 2007, the publication list including academic journals, book chapters, and conference proceedings stood at more than 4000 [Emrouznejad et al., 2008]. Additional bibliographic listings are available at http://www.deazone.com/.

1.5.2 Numerical and Graphical Example

To illustrate how DEA works, we consider a simple example consisting of five DMUs with the same scale (labeled A to E), each consuming one single input to produce one single output (see Table 1.3).

Table 1.3 Single input and single output case

DMU	Input	Output	Efficiency
A	2	1	0.5
B	3	3	1
C	5	2	0.4
D	6	5	0.833
E	7	4	0.571

Based on the input and output values shown in the second and the third column of Table 1.3, we can calculate the corresponding efficiency score for each DMU according to the expression (1-1). The results are shown in the last column of Table 1.3. Using this measure, we can identify *B* as the most efficient DMU while *C* is the least efficient.

Let's represent these data as in Figure 1.7 by plotting the input value of each DMU on the horizontal axis and the output value on the vertical axis. The slope of the line connecting each point to the origin thus indicates the efficiency score of the corresponding DMU, and the line with the maximum slope, which is from the origin through *B* shown in Figure 1.7, is the efficient production frontier for all DMUs being analyzed. The DMUs that are on this line correspond to efficient units, while those below the frontier are inefficient. The area between the frontier and the positive horizontal semi-axis is called the *production possibility set*, and all the DMUs are – in mathematical parlance – 'enveloped' within this set.

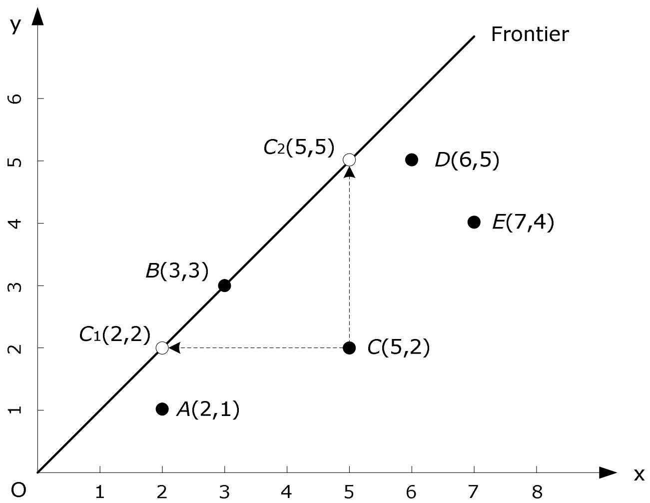


Figure 1.7 Graphic representation of the efficiency production frontier

In this case, in contrast with the best performer *B*, all the others are inefficient. We can then measure the efficiency of all the DMUs relative to *B* by

$$0 \leq \frac{\text{Efficiency of each DMU}}{\text{Efficiency of DMU } B} \leq 1 \tag{1-2}$$

and arrange them in the following order: $1 = B > D > E > A > C = 0.4$. Thus, DMU B obtains the highest relative efficiency of 1, which differs in meaning from the one in Table 1.3 since it is 'unit invariant', and the worst-performer C, attains only 40% of B's efficiency.

Notice that the efficient production frontier also provides some indications for improving the performance of inefficient units. Taking DMU C in Figure 1.7 as an example, its efficiency can be improved in several ways. One is achieved by reducing the input to C_1 with coordinates (2,2) on the frontier. Another is achieved by raising the output up to C_2 (5,5). In fact, any point on the line segment C_1C_2 offers a chance to make DMU C efficient in a manner which assumes that the input should not be increased and the output should not be decreased.

The graphical representation provides a visual description of the fundamental mechanisms of DEA, and it is useful in simple one or two dimensional examples. However, it becomes difficult when higher dimensions are considered. The normal approach is therefore to employ a *linear programming* formulation of DEA so as to estimate the efficient production frontier of the set of DMUs, and also to measure the relative efficiency of each DMU under consideration.

1.5.3 The DEA models

As a frontier analysis technique, DEA applies mathematical optimization techniques to estimate the relations between multiple inputs and multiple outputs related to a homogeneous set of DMUs. During these years, a number of different formulations have been proposed in the DEA context, the best-known of which is probably the Charnes–Cooper–Rhodes (CCR) model. Specifically, consider an n-DMUs set, each consuming m different inputs to produce s different outputs. The relative efficiency of a DMU is defined as the ratio of its total weighted output to its total weighted input, subjected to (*s.t.*) lie between zero and the unity. Mathematically, the efficiency score of a particular DMU_0, denoted as E_0, is obtained by solving the following constrained optimization problem:

$$\max E_0 = \frac{\sum_{r=1}^{s} \mu_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$$

$$s.t. \quad \frac{\sum_{r=1}^{s} \mu_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad j = 1, \cdots, n \tag{1-3}$$

$$\mu_r, v_i \geq 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m$$

where y_{rj} and x_{ij} are the rth output and ith input, respectively of the jth DMU, μ_r is the weight given to output r, and v_i is the weight given to input i. This fractional program is computed separately for each DMU to determine its optimal input and output weights. In other words, the weights in the objective function are chosen automatically from the model with the purpose of maximizing the value of DMU_0's efficiency ratio and also respecting the less than unity constraint for all the DMUs. Meanwhile, all the weights are required to be non-negative. This condition is sometimes replaced by using a small non-Archimedean number $\varepsilon > 0$ for restricting the model to assign a weight of zero to unfavorable factors [Charnes & Cooper, 1984].

To simplify the calculation and to avoid an infinite number of solutions[1], the above fractional program can be converted into a linear program, which is known as *the multiplier form* of the problem:

$$\max E_0 = \sum_{r=1}^{s} u_r y_{r0}$$

$$s.t. \quad \sum_{i=1}^{m} v_i x_{i0} = 1, \tag{1-4}$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \cdots, n$$

$$u_r, v_i \geq 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m$$

The transformation is completed by constraining the efficiency ratio denominator (i.e., the weighted sum of inputs) in (1-3) to a value of one. Thus the objective function consists of the maximization of the weighted sum of outputs. It is also named the *input-oriented* CCR model. Analogously, the *output-oriented* one can

[1] If (μ^*, v^*) is an optimal solution, then $(a\mu^*, av^*)$ is also optimal for $a > 0$ [Cooper et al., 2004].

be derived by requiring the weighted sum of outputs to be one and then minimizing the weighted sum of inputs (see also Cooper et al. (2004)). In general, a DMU is considered to be efficient if it obtains an efficiency score of one in (1-4), whereas a score less than one implies that it is inefficient.

Furthermore, using the duality in linear programming, we can derive an equivalent *envelopment form* of the above problem, which can be formulated as follows:

$$
\begin{aligned}
\min \quad & \theta_0 \\
s.t. \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_0 x_{i0}, \quad i = 1,\cdots,m \\
& \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}, \quad r = 1,\cdots,s \\
& \lambda_j \ge 0, \quad j = 1,\cdots,n
\end{aligned}
\tag{1-5}
$$

where θ_0 is the uniform proportional reduction in the DMU_0's inputs. Its minimum amount is known as the DMU_0's efficiency score, which also equals to E_0 calculated in (1-4). Moreover, λ_j ($j = 1,\cdots,n$) is the dual weight given to the jth DMU's inputs and outputs in constructing for DMU_0 a *hypothetical composite unit* (HCU) which lies on the efficient production frontier and produces at least as much of each output as DMU_0 does, yet meanwhile consumes no more than the proportion θ_0 of its inputs. For each feasible solution (θ_0, λ) to problem (1-5), the *slack* variables s_i^- ($i = 1,\cdots,m$) and s_r^+ ($r = 1,\cdots,s$) can be defined, which represent respectively the quantity of input i used in excess by DMU_0 and the quantity of output r produced in shortage by DMU_0 with respect to the HCU:

$$
\begin{aligned}
s_i^- &= \theta_0 x_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_j, \quad i = 1,\cdots,m \\
s_r^+ &= \sum_{j=1}^{n} y_{rj}\lambda_j - y_{r0}, \quad r = 1,\cdots,s
\end{aligned}
\tag{1-6}
$$

Hence, if the value of θ_0 equals one and the value of all the slack variables equals zero, it means that no input reduction is needed for DMU_0 to produce its output. In other words, it is efficient and its input-output combination lies on the frontier. If $\theta_0 < 1$, DMU_0 is said to be *technically inefficient*, and it lies inside the frontier. It means that in order to obtain the same outputs, the inputs used could be simultaneously reduced with the proportion $1-\theta_0$. In addition, if $\theta_0 = 1$, but some slack variables are different from zero, DMU_0 then presents a *mix inefficiency* since, keeping the same output level, it could reduce the consumption of some inputs without causing an increase in other inputs used. No matter which kind of inefficiency the DMU under consideration belongs to, those DMUs that contribute to the construction of the HCU (with a non-zero value of λ in the optimal solution of the dual model) make up the reference set for it and can be treated as its benchmarks. Moreover, based on the value of λ, the relative importance of a DMU within the reference set can be identified, and the target values for the inefficient DMU can be determined as follows[2] [El-Mahgary et al., 1995]:

$$\begin{aligned} x_{i0}^{target} &= \sum_{j=1}^{n} x_{ij}\lambda_j = \theta_0 x_{i0} - s_i^-, \quad i = 1,\cdots,m \\ y_{r0}^{target} &= \sum_{j=1}^{n} y_{rj}\lambda_j = y_{r0} + s_r^+, \quad r = 1,\cdots,s \end{aligned} \tag{1-7}$$

As indicated in the beginning of this section, DEA is regarded as a body of concepts and methodologies that has evolved since the seminal work of Charnes, Cooper & Rhodes (1978). Apart from the CCR model introduced in this section, a large number of other DEA models have also been successfully proposed and widely investigated during the last decades. Some of them are the Banker-Charnes-Cooper (BCC) model, the additive model, the slacks-based measure of efficiency (SBM), the multiplicative model, and so on [Cooper et al., 2000]. Different DEA models have their own specific characteristics. For instance, the BCC model takes into account the most productive scale size while simultaneously identifying technical inefficiency [Banker et al., 1984], such that the assumption of *variable returns to scale* (VRS) rather than *constant returns to scale* (CRS) as in the CCR model is utilized, which produces results reflecting the contribution of factors such as the size of the DMU (see Banker et al. (2004) for discussions on returns-to-scale in DEA). The additive model is an extension of the BCC model and has VRS, but its main difference is that both input and

[2] The target values in (1-7) emphasize input reduction since the input-oriented model is considered. When output enhancement is the main concern, the output-oriented model should be adopted. For more information, we refer to [Cooper et al., 2004].

output orientations are combined in a single model by seeking to project the DMU values to the most distant point of efficiency on the frontier using only non-zero slacks. The SBM further extended the additive model 'to give scalar measures from 0 to 1 that identify all of the inefficiencies that the model can identify' [Cooper et al., 2000]. Finally, the multiplicative model uses piecewise log-linear or piecewise Cobb-Douglas envelopment instead of the traditional linear piecewise surface and focuses on the impact of the weighting factors. In general, selecting which DEA model to use is to a great extent case-based. In this dissertation research, the CCR model is the main focus to highlight the fundamental mechanisms that have propelled the application of DEA. More importantly, the CRS assumption in the CCR model reflects the linear character between inputs and outputs when either one is changed. It is particularly suitable in our road safety benchmarking study, which guarantees the comparisons of road safety performance and development between countries on the same presupposition.

1.5.4 Strengths and limitations of DEA

DEA has proven valuable as a non-parametric optimization technique for measuring the relative efficiency of a homogeneous set of DMUs by allowing direct peer comparisons on the basis of multiple inputs and multiple outputs through a diverse range of models. As one of the powerful benchmarking tools, DEA has received significant attention in the last decades due to its prominent advantages over other traditional methods, such as the ratio analysis, the OLS techniques, and so on, as mentioned in Section 1.2.2.

First of all, DEA provides a new way of obtaining empirical estimates of input-output relations by constructing an efficient production frontier based only on the best performers within the observations. Thus, there is no need to make any assumptions about the functional form of the frontier which is often complex or even unknown in the real world situation [Charnes et al., 1994].

Second, DEA is capable of using multiple inputs and multiple outputs simultaneously, which is superior to the simple ratio analysis that provides only partial measures of the multiple input-output relations and thus often leads up to misclassifications and incorrect judgments [Lewin et al., 1982].

Moreover, the inputs and outputs used in the model can be expressed in different units of measurement. In other words, the preliminary normalization (e.g., standardization) of raw data is not required, which is particularly convenient from a practical point of view and eliminates the sensitivity of the results with respect to the specific normalization scheme that is used [Organization for Economic Co-operation and Development, 2008].

Fourth, an a priori knowledge concerning the input and output weights is not necessary (although sometimes preferable) in DEA. Moreover, by applying the multiplier form of the model, each DMU obtains its own best possible input and output weights, which is objective in nature and also different from most of the other weighting schemes that assign the same input and output weights for all the DMUs, such as AHP. The flexibility enjoyed by the DMUs in choosing their own input and output weights represents an undisputed advantage, in that if a DMU turns out to be inefficient based on the most favorable set of weights, its inefficiency cannot be traced back to an inappropriate evaluation process [Vercellis, 2009].

In addition, DEA assesses the relative efficiency of a particular unit by comparing it against all others, and the final efficiency score is measured with respect to the best observed performance, which is different from other techniques such as the OLS and the SFA that are based on either the average observed or some predetermined performance [Hayashi, 2000; Kumbhakar & Lovell, 2000].

Last but not least, by distinguishing between efficient units and inefficient units using the dual form of the model, DEA possesses the ability to analyze the sources of inefficiency and further determine the potential improvement for those inefficient units by indicating specific benchmarks and assigning practical targets for them, which mostly attracts analysts and decision makers, and results in the widespread application of this technique [Amirteimoori et al., 2005; Hermans et al., 2009b; Yang et al., 2009].

On the other hand, the same characteristics that make DEA a powerful tool can also create problems. An analyst should keep these limitations in mind when deciding whether or not to use DEA:

Above all, as an extreme point technique, DEA is susceptible to 'noise' (even if it is symmetrical with zero mean) such as measurement error, and it can be similarly affected by an outlier impact especially when the number of DMUs increases dramatically [Golany & Roll, 1989]. Some outliers may exist on the production frontier thus changing the envelopment surface used for peer comparisons and affecting which ones are rated as efficient.

There is an equally strong argument presented concerning the selection of inputs and outputs related to the DMUs as well as their quantity. The results of DEA are sensitive to the inputs and outputs utilized, and the number of efficient DMUs on the frontier tends to increase with the number of inputs and outputs. As a result, the discriminating power of the model is lost [Cooper et al., 2000].

The other widely debated feature of DEA is the assignment of weights to the various factors. The flexibility in selecting the weights in DEA is often presented as advantageous in its applications since a priori specification of the weights is not required and each DMU is evaluated in its best possible light. However, it also makes the comparison among DMUs on a common base impossible. Moreover, an unreasonable weight scheme could happen in which some DMUs would heavily weigh a few favorable inputs and outputs and completely ignore others in order to achieve a high relative efficiency score [Wong & Beasley, 1990; Allen et al., 1997; Thanassoulis et al., 2004]. Therefore, given a DMU that obtains an efficient score of one, it is important to determine whether its efficiency value should be attributed to an actual high-level performance or simply to an optimal selection of the weight structure.

In addition, the non-parametric nature of DEA means that it does not allow the application of inferential statistics and traditional mechanisms such as hypothesis testing, which is actually the focus of ongoing DEA research [Zhu & Cook, 2007; Bogetoft & Otto, 2011].

Although the limitations indicated above may be inimical to the successful applications of DEA, it should be noted that a better understanding of their threat and possible impact also provides the directions for future investigation.

1.5.5 Model extensions for this research

Due to its prominent advantages, DEA is considered as an appropriate tool for our road safety benchmarking studies. However, to successfully apply DEA in this dissertation research, some methodological extensions have to be explored to handle some of the model limitations mentioned in the previous section and to answer the specific research questions listed in Section 1.4 as well.

1.5.5.1. DEA-based road safety model (Model 1)

In the basic DEA model, the definition of the best practices relies on the assumption that inputs have to be minimized and outputs have to be maximized. However, to use DEA for road safety evaluation, we want the outputs, e.g., the number of road fatalities, to be as low as possible with respect to the level of exposure to risk, such as distance travelled. As a result, a DEA-based road safety model is developed (see Chapter 2), in which the frontier DMUs, or the road safety best-performing countries are those with minimum output levels given input levels, and other countries' efficiency is then measured relative to this frontier.

1.5.5.2. Cross-efficiency method (Model 2)

As indicated in Section 1.5.4, DEA possesses an attractive feature that each DMU is allowed to select its own most favorable input and output weights for calculating its best efficiency score, rather than the same weights for all the DMUs. However, this flexibility in selecting the weights makes the comparison among DMUs on a common base impossible. To overcome this difficulty, a cross-efficiency method [Sexton et al., 1986] was developed with the main idea of using DEA in a peer evaluation instead of a self-evaluation mode (see Chapter 2). In this way, the results can be used to identify the best overall performers and to effectively rank all DMUs.

1.5.5.3. Categorical DEA model (Model 3)

As a remarkable benchmarking approach, DEA owns the capability of indicating a specific reference set for those inefficient DMUs and determining their potential improvement, or target, which is particularly suitable for answering our second research question. However, the traditional benchmarking analysis also has some limitations – an inefficient DMU and its corresponding reference set may not be inherently similar in their practices, or the benchmarks are probably too far away for the inefficient DMU to learn from – which means that the resulted target may be very unrealistic and not attainable for this inefficient DMU. To solve this problem, clustering analysis is adopted to first cluster DMUs into a number of groups, and the best performers in a particular cluster, derived from a categorical DEA model, are then utilized as benchmarks for other DMUs in the same cluster (see Chapter 2).

1.5.5.4. DEA-based Malmquist index (Model 4)

Apart from assessing the road safety performance of countries at one specific point of time, research on time series data collected at regular intervals is also valuable in gaining a clear understanding of trends and expected progress towards postulated targets and benchmarks' performance, which is our third research question. For this purpose, a DEA-based Malmquist productivity index, initially proposed by Malmquist (1953), can be applied. It has proven to be a proper tool for measuring the total factor productivity change of a DMU, in that it reflects progress (or regress) in efficiency along with progress (or regress) of the frontier technology over time [Chen & Ali, 2004; Yörük & Zaim, 2005; Greer, 2008]. The calculation of this index for the road safety context will be illustrated in Chapter 3.

1.5.5.5. Weight restrictions in DEA (Model 5)

As a widely debated feature of DEA, an unreasonable weight scheme could happen due to the flexible allocation of input and output weights. To separate the DMUs that are really efficient from those whose efficiency score largely depends on the selected weights, apart from the cross-efficiency method, we may also impose different types of restrictions on the value of the weights to be associated with inputs and outputs. In doing so, a priori knowledge or requirements on the weights and also the value judgments from decision makers or experts can be incorporated. See Chapter 4 and Chapter 6 for the detailed discussion on weight restrictions.

1.5.5.6. DEA-based composite index (Model 6)

The basic DEA model assesses the relative efficiency of a set of DMUs on the basis of multiple inputs and multiple outputs. However, to use DEA for composite index construction as in this research, i.e., aggregating a set of individual indicators into one overall index, it means that only inputs or outputs of the DMUs will be taken into account in the model. As noted by Adolphson et al. (1991), it is possible to adopt a broader perspective, in which DEA is also appropriate for comparing any set of homogeneous units on multiple dimensions. Based on this perspective, the DEA-based composite index model is realized (See Chapter 6), which is also known as the 'benefit of the doubt' approach [Cherchye et al., 2007a].

1.5.5.7. Multiple layer DEA model (Model 7)

One significant limitation of DEA indicated in Section 1.5.4 is that a large number of DMUs will obtain an efficiency score of one when the number of inputs and outputs used in the model is too large relative to the number of DMUs. On the other hand, as performance management becomes more and more complex, there are a great number of performance evaluation activities which not only need to be represented by a continuously increasing number of performance indicators, but these indicators might also belong to different categories and further be linked to one another constituting a multilayer hierarchical structure (corresponding to our sixth research question). In these cases, simply treating all the inputs and outputs to be in the same layer obviously ignores the information on the hierarchical structure of the indicators, and further leads up to weak discriminating power and unrealistic weight allocations. To this end, a multiple layer DEA model (MLDEA) is proposed in this research and applied to the road safety composite index construction (See Chapter 6).

1.5.5.8. Interval DEA model (Model 8)

As a 'data-oriented' technique, the applicability of DEA relies firstly on the availability of data. In other words, a complete data set with crisp positive values is commonly the prerequisite of the evaluation. However, in many applications, the efficiency evaluation of the DMUs has to take into account missing values for some inputs and outputs (research question seven), which to a great extent restricts researchers from performing the basic DEA models. One of the possible solutions is to use interval DEA [Despotis & Smirlis, 2002; Smirlis et al. (2006); Cherchye et al., 2011], in which the missing values are replaced by approximations in the form of intervals in which the true values are believed to lie. The bounds of the intervals, depending on the application, can be estimated by using statistical or experiential techniques. The model provides for the DMUs with missing values an upper and a lower bound of their efficiency score corresponding to their most favorable and unfavorable option, respectively (See Chapter 7).

1.5.5.9. Imprecise and Fuzzy DEA model (Models 9 & 10)

Apart from data availability, the quality of data is also vital to the successful application of DEA. However, measurable or quantitative data are sometimes inadequate or even inappropriate to represent real world situations due to the complexity and uncertainty of the reality. Therefore, it is essential to take into account the presence of qualitative data when making a decision on the performance of a DMU, which corresponds to our eighth research question. Under these circumstances, the basic DEA model is out of its capability to derive a satisfactory solution. Generally, two strategies have appeared in the literature to the treatment of qualitative data within the DEA framework. One is to reflect the rank position of each DMU with respect to each ordinal factor by setting corresponding constraints, which is collectively referred to as imprecise DEA [Cook & Zhu, 2006]. The other strategy is to deal with the natural uncertainty inherent to some production processes by means of fuzzy mathematical programming, which leads up to the so-called fuzzy DEA [Guo, 2009]. In this research, both strategies are investigated to model qualitative data for the road safety benchmarking study (See Chapter 8).

1.6 Structure of the Dissertation

This thesis consists of nine chapters, with the structure illustrated in Table 1.4. This first chapter provided a general introduction to the dissertation research.

Chapters 2 to 4 constitute the first thematic part in which the road safety final outcomes are investigated for road safety product benchmarking. Chapters 5 to 8 form together the second thematic part, which focuses on analyzing road safety performance indicators within the road safety programme benchmarking context. The dissertation ends with final conclusions and guidelines for future research in Chapter 9.

Table 1.4 Structure of the dissertation

Chapter 1 General Introduction			
Part I Benchmarking Road Safety Development: Evidence from Final Outcomes	**Chapter 2** Road Safety Risk Evaluation and Target Setting on Fatalities	RQs 1 & 2	Models 1, 2, & 3
	Chapter 3 Road Safety Development in Europe: A Decade of Changes (2000-2009)	RQ 3	Models 1 & 4
	Chapter 4 Serious Injuries: An Additional Indicator for Road Safety Evaluation	RQ 4	Models 1 & 5
Part II Towards a Composite Road Safety Performance Index	**Chapter 5** Development of Safety Performance Indicators and Data Processing	RQ 5	
	Chapter 6 Construction of a Composite Index (I): Hierarchical Structure Assessment	RQ 6	Models 2, 4, 6, & 7
	Chapter 7 Construction of a Composite Index (II): Taking Interval Data into Account	RQ 7	Models 6, 7, & 8
	Chapter 8 Construction of a Composite Index (III): Modeling Qualitative Data	RQ 8	Models 6, 7, 9, & 10
Chapter 9 Final Conclusions and Future Research			

Chapter 2 evaluates the overall road safety risk of a set of countries by simultaneously taking the different measures of exposure to risk into account, and identifies specific benchmarks and assigns practical targets for those underperforming countries in terms of their number of road fatalities. This chapter thereby corresponds to the first two research questions of this dissertation. In doing so, DEA and its three model extensions (i.e., the DEA-based road safety model, the cross-efficiency method, and the categorical DEA model) are investigated.

Chapter 3 focuses on the third research question of this dissertation, i.e., how to evaluate the road safety performance change of countries over time. In doing so, we not only focus on the evolution in the number of road fatalities within a given period, but also take the change in exposure in the same period into account. The DEA-based road safety model and the Malmquist productivity index are employed to undertake the assessment.

Chapter 4 discusses the possibility of including the number of serious injuries as an additional indicator of road safety final outcome to perform road safety product benchmarking and further illuminates its impact on the countries' ranking. It thereby corresponds to the fourth research question of this dissertation. In doing so, different types of weight restrictions are formulated in the DEA-based road safety model to indicate the relationship between road fatalities and serious injuries.

Chapter 5 identifies the current available national safety performance indicators that could be used for inter-national programme benchmarking of road safety, thereby corresponding to the fifth research question of this dissertation. Given the various categories of risk factors to consider and the idea of representing each risk factor with a number of safety performance indicators, the hierarchical structure of the indicators is established. Moreover, outliers in the data set are examined, and missing values are imputed. The complete data set provides us with the basis for the following road safety performance index research.

Chapter 6 elaborates on the use of a DEA model for composite index construction, especially when the hierarchical structure of the indicators is taken into account, which therefore answers the sixth research question of this dissertation. The proposed multiple layer DEA model is applied to combine the hierarchical indicators developed in the previous chapter into a composite road safety performance index. Useful insights are gained from benchmarking analyses enabling policymakers to prioritize their actions to improve the level of road safety.

Chapter 7 investigates the influence of missing data on the final index score. In doing so, missing data are firstly replaced by approximations in the form of intervals deduced from multiple imputation in which the true values are believed to lie. An interval MLDEA-based CI model is thereafter developed and applied to provide for each country an upper and a lower bound of its index score corresponding to its most favorable and unfavorable option, respectively. This chapter therefore deals with the seventh research question of this dissertation.

Chapter 8 focuses on the last research question of this dissertation, which is to model qualitative data in the context of composite index construction. Two strategies, i.e., an imprecise DEA-based CI model and a fuzzy DEA-based CI model are thereby investigated. The models are demonstrated by taking the qualitative alcohol indicator developed in Chapter 5 into account. Furthermore, by integrating fuzzy logic into the MLDEA-based CI model proposed in Chapter 6, we obtain a fuzzy MLDEA-based CI model, which is capable of combining all the hierarchical indicators (with both quantitative and qualitative data) into a composite road safety performance index.

Finally, Chapter 9 summarizes the main conclusions from this dissertation research. Moreover, directions for future research are provided.

Part I Benchmarking Road Safety Development: Evidence from Final Outcomes

Introduction to Part I

In most of the current road safety benchmarking studies, crash data such as the number of fatalities – traditionally seen as road safety final outcomes or road traffic by-products – are inevitably investigated for all the countries under consideration. Their percentage change over the given period and the concept of risk, which is defined as the ratio of the final outcomes and some measure of exposure, are often used to make countries comparable. Evaluation and comparison based on these figures are accordingly also named as road safety product benchmarking, which enables countries not only to better understand their own relative road safety situation in terms of final outcomes, but more importantly, to learn from those better-performing countries in developing their own road safety policy, such as setting practical targets on their final outcomes.

However, in computing the level of risk for a country, different exposure information can be used (e.g., the population size, the number of registered vehicles, and the distance travelled), and different evaluation results or ranking positions are normally obtained based on different risk indicators. So far, there is no consensus about which one is the most appropriate indication, and research on their combination is also quasi non-existing.

Moreover, while road safety product benchmarking based on risk assessment enables us to find a specific reference set for those underperforming countries and further determine their potential improvement, e.g., the target number of fatalities, there are still some limitations involved in its practice such as that a underperforming country and its corresponding reference set may not be inherently similar, or the benchmarks are probably too far away for the underperforming country to learn from.

Furthermore, to compare the development of road safety between countries, the percentage change in the number of people killed on the road is often the main indicator. However, simply considering the change in the final outcome may not correctly reflect the real development of road safety because the transport circumstances of a country having an impact on the final outcome, also changes over time.

In addition, from the view of road safety final outcomes, most of the current road safety product benchmarking studies focus entirely on fatalities, which however, represent only one measure of the magnitude of the road safety problem. Consequently, it is desirable to extend the inter-national comparisons of road safety by taking a larger picture of the impact of road crashes into account, such as serious injuries.

In the first thematic part of this thesis, four research questions related to road safety product benchmarking are thereby formulated as below, and answers are given in the following three chapters.

RQ1: How to obtain an overall picture of a country's road safety performance when different final outcome risk indicators are considered?

RQ2: How to set practical targets for similar underperforming countries in terms of the number of road fatalities?

RQ3: How to assess road safety performance change over time by taking both final outcome evolution and exposure change into account?

RQ4: What is the impact of utilizing other final outcomes, e.g., serious injuries, in addition to fatalities for road safety product benchmarking?

Chapter 2 Road Safety Risk Evaluation and Target Setting on Fatalities[3]

This chapter evaluates the overall road safety risk of a set of countries by simultaneously taking the different measures of exposure to risk into account, and further identifies specific benchmarks and assigns practical targets for those underperforming countries in terms of their number of road fatalities. This chapter thereby corresponds to the first two research questions of this dissertation.

2.1 Introduction

Currently, the road safety situation of a country is mostly evaluated by means of crash data such as the number of road fatalities. However, the absolute numbers are not directly comparable between countries. In other words, the scale of different countries has to be considered when making comparisons. Therefore, the concept of risk, which is defined as the ratio of road safety outcomes and some measure of exposure, is often used in the context of benchmarking [European Transport Safety Council, 2003]. In this respect, the population size, the number of registered vehicles, and the distance travelled are the three most frequently used measures of exposure to risk [International Traffic Safety Data and Analysis Group, 2011]. Nevertheless, there has been considerable debate in the past about which one is the most appropriate indicator of exposure, because they describe risk from different points of view and are not consistent in most cases. In other words, countries may have different evaluation results or ranking positions using different exposure information, which to a great extent baffles decision makers in assessing their relative performance. Therefore, it would be desirable if all these three risk

[3] Related research has been published in: Shen, Y., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K.. Road safety risk evaluation and target setting using data envelopment analysis and its extensions, Accident Analysis & Prevention, doi: 10.1016/j.aap. 2012.02.020.

indicators could be considered together in order to make comparisons of performance between countries. From the target setting point of view, however, numbers rather than rates are much more preferred since a declining rate such as the fatalities per distance travelled may conceal an increase in the raw number of fatalities [European Road Safety Observatory, 2006]. Consequently, an analytical research tool that can represent an overall perspective on a country's road safety situation (in ratios, which make countries comparable), and also provide improvement potential for those underperforming countries (by numbers), is required.

In this research, data envelopment analysis (DEA) and its several extensions, including the DEA-based road safety model, the cross-efficiency method, and the categorical DEA model are investigated and applied to evaluate the overall road safety performance of a country by simultaneously taking the aforementioned three aspects of exposure to risk into account, and to further assess whether the road safety outcomes registered in a country correspond to the ideal numbers that can be expected based on the level of exposure. The analysis, which is based on the combination of these model extensions, provides interesting insights and valuable recommendations for road safety policymakers in identifying the 'efficiency' of their current operations (i.e., an efficient transformation of input or exposure into output or road safety outcomes) and in suggesting useful benchmarks and practical targets for improvement.

The remaining of this chapter is structured as follows. Section 2.2 introduces the three main risk indicators for road safety evaluation, and the idea of setting quantitative road safety targets is presented in Section 2.3. Section 2.4 specifies the relevant extensions of the basic DEA model for this study. The application of the methodology to the road safety risk evaluation and target setting is illustrated in Section 2.5 and the corresponding results are discussed subsequently. The chapter ends with the main conclusions in Section 2.6.

2.2 Risk Indicators

Reduction of road traffic crash risk and consequent damage, injury, and death is the key objective of policy concerning road safety. In order to obtain numerically reliable estimates of risk, recorded numbers of fatalities are usually related to some measure of exposure, which is currently the main form of risk assessment in road transport between countries[4] [European Transport Safety Council, 2003].

Concerning exposure to risk, population data are most commonly used since they are readily available in most countries. The corresponding risk indicator, i.e., the number of fatalities per million inhabitants, is known as *the mortality rate* and regarded as an important criterion for road safety evaluation since it permits comparisons with other causes of death such as heart disease. However, for the comparison of traffic risks this indicator has the disadvantage of leaving the level of motorization out of account. Accordingly, an estimation of exposure to risk in terms of traffic volume is introduced representing *the fatality risk*, which is defined as the number of fatalities per distance travelled (e.g., fatalities per 10 billion passenger-kilometres (pkm) travelled). This risk indicator has traditionally been favored by road transport authorities as it implicitly discounts fatality rates if travel is increased. However, the definition of this exposure measure differs widely across countries, and only a limited number of countries collect data on this exposure measure. As a result, a third risk indicator – defined as the number of fatalities per million registered vehicles, which is also called *the fatality rate* – is often used as a substitution, although it differs in that the annual distance travelled is unknown. In addition, there are still some other variables which can be used as measures of exposure, such as road length, fuel consumption, the number of driving license holders, and so on. For more information, we refer to [Yannis et al., 2008].

In the European Commission's report on EU Energy and Transport in Figures [European Commission, 2010b], 2008 data related to the above three risk indicators are collected for the 27 EU Member States, which are Austria (AT), Belgium (BE), Bulgaria (BG), Cyprus (CY), Czech Republic (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IE), Italy (IT), Latvia (LV), Lithuania (LT), Luxembourg (LU), Malta (MT), the Netherlands (NL), Romania (RO), Poland (PL), Portugal (PT),

[4] Fatalities are used in most road safety analyses not because they are the only interest but mainly because there is no reliable reporting of the number of crashes and the range of injury severities. Even their definition varies greatly among countries.

Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE), and United Kingdom (UK). Country rankings in decreasing order of safety are indicated in Table 2.1.

Table 2.1 Rankings of the 27 EU countries based on the three road safety risk indicators in 2008

Fatalities per million inhabitants		Fatalities per 10 billion pkm		Fatalities per million passenger cars	
MT	36	UK	39	MT	66
NL	41	SE	40	UK	91
SE	43	NL	45	NL	91
UK	43	LU	51	SE	93
DE	55	DE	51	LU	108
IE	63	FI	53	DE	109
FI	65	IE	56	FI	131
ES	68	FR	58	IT	132
FR	69	IT	59	FR	137
LU	72	MT	68	ES	141
DK	74	DK	75	IE	144
EU-27	78	EU-27	80	AT	159
IT	79	BE	84	EU-27	168
AT	81	SI	85	BE	185
PT	83	ES	89	CY	192
BE	88	AT	91	DK	195
EE	98	PT	99	PT	201
HU	99	EE	124	SI	208
SK	103	LT	129	EE	245
CZ	103	CY	139	CZ	247
CY	103	CZ	142	LT	306
SI	106	EL	147	EL	317
EL	138	LV	181	HU	328
BG	139	PL	196	LV	344
LV	139	SK	206	PL	355
RO	142	HU	230	SK	375
PL	143	BG	241	BG	477
LT	148	RO	420	RO	809

Notes:

Fatalities: number of persons who were recorded as dying immediately or within 30 days from injuries sustained in a collision.

Inhabitants: sum of the population at 1 January 2008 and 1 January 2009 divided by two.

pkm: passenger-kilometres of cars plus passenger-kilometres of motorised two-wheelers.

Passenger cars: sum of the stock of vehicles for 2007 and 2008 divided by two.

Source: European Commission (2010b)

These three risk indicators describe the relative performance in road safety of the 27 EU countries from different perspectives. However, their rankings also vary from one indicator to another. For instance, United Kingdom ranks first with respect to the fatalities per 10 billion pkm, but not with respect to the other two exposure measures. In fact, this is the case for all countries. Such kind of inconsistencies baffles the decision makers in identifying the best-performing countries and in deciding the extent to which those underperforming ones should improve. Consequently, obtaining an overall picture of a country's road safety performance for cross-country comparison is valuable.

2.3 Target Setting

If we argue that risk analysis has the potential to make a powerful contribution to the development of effective strategies and programmes for crash prevention and casualty reduction, then the setting of challenging yet achievable quantitative road safety targets (usually expressed in terms of final outcomes, e.g., reduction in the number of fatalities) serves as a significant catalyst that motivates the whole range of stakeholders (from individuals who use the roads in different ways to government agencies at all levels) to support such strategies and programmes in order to achieve the safer use of roads. The value of setting targets to reduce road fatalities and casualties and thereby improve road safety performance has been widely recognized, see also [Elvik, 2001; Wong, 2006; Allsop et al., 2011]. An increasing number of countries are implementing long term road safety strategies towards their reduction or eventual elimination (e.g., the Swedish Vision Zero [Organization for Economic Co-operation and Development/International Transport Forum, 2008a]) within a framework of quantitative road safety targets. A range of targets in current use of a number of EU Member States is described in Table 2.2.

Table 2.2 Quantitative road safety targets in Europe

Country	Base year	Target year	Targets on road fatalities
BE	Mean of 1998-2000	2015	-33% (max 500 fatalities)
DK	1998	2012	-40%
EE	2002	2015	-55%
EL	2000	2015	-40%
FI	2000	2025	-75% (less than 100 fatalities)
HU	2001	2015	-50%
MT	2004	2014	-40%
NL	Mean of 2000-2002	2020	Less than 580 fatalities
RO	2002	2012	-50%
EU-27	2010	2020	-50%

Source: Elvik (2003); *Organization for Economic Co-operation and Development /International Transport Forum (2006)*; *European Commission (2010a)*

No matter whether the target is expressed in percentage reductions or in absolute numbers, it represents the desired road safety results that a country wishes to achieve over a given timeframe. In practice, setting a challenging yet achievable quantitative target, however, is by no means easy. It needs to be ambitious in order to render all the stakeholders to come together and be motivated to share their responsibility in achieving common safety goals. Meanwhile, it should also be realistic so as to keep and strengthen this motivation during the whole target period. In doing so, many factors have to be taken into account, such as the economic status of a country, the level of ambition and commitment, the potential of different measures, the available resources, and so on. In the current research on target setting, such as [Organization for Economic Co-operation and Development, 2002], estimates of what is likely to be achievable are mostly based upon information about the current road safety situation of a country and its past evolution. In doing so, a reasonable assumption about the future is required, which, however, is to a great extent untraceable due to the complexity and uncertainty of the reality. In this study, an alternative way for target setting is introduced, in which each country is allowed to learn from other countries' best performance. More specifically, since the achievements that have already been captured by those best-performing countries provide valuable directions for the underperforming ones to go forward, the target value of one country can then be determined by its benchmarks. This objective is realized by using the technique of data envelopment analysis.

2.4 Methodology

Data envelopment analysis, as introduced in Section 1.5, is a powerful benchmarking tool with some prominent advantages over other traditional methods, which has resulted in the widespread application of this technique to a large number of benchmarking studies. However, to use DEA in this road safety evaluation and target setting research, some model considerations are still needed, which will be elaborated in the following sections.

2.4.1 DEA-based road safety model

In the basic DEA model, the definition of the best practices relies on the assumption that inputs have to be minimized and outputs have to be maximized (such as in the economics field). However, to use DEA for road safety evaluation, we want the output, i.e., the number of road fatalities to be as low as possible with respect to the level of exposure to risk. Therefore, the DEA frontier DMUs, or the road safety best-performing countries are those with minimum output levels given input levels, and other countries' efficiency is then measured relative to this frontier. Graphically, consider the fatality risk (i.e., the number of fatalities per passenger-kilometres travelled) of two countries P and Q as in Figure 2.1.

According to the DEA-based road safety principle that for a given amount of pkm travelled, countries having a lower number of fatalities are the efficient ones, we can thus identify that country Q is efficient. Thereby, the efficiency production frontier (*F*) is the ray extending from the origin through Q, and the area above this frontier constitutes the production possibilities set, i.e., the set of feasible activities, in which country P is located. Hence, P is inefficient, and its efficiency score can be computed as: AB/AP<1. Therefore, without any change of the exposure level, the number of fatalities in country P should be proportionally reduced by 1-AB/AP=BP/AP to become efficient, and thus point B could be treated as its hypothetical composite unit (HCU).

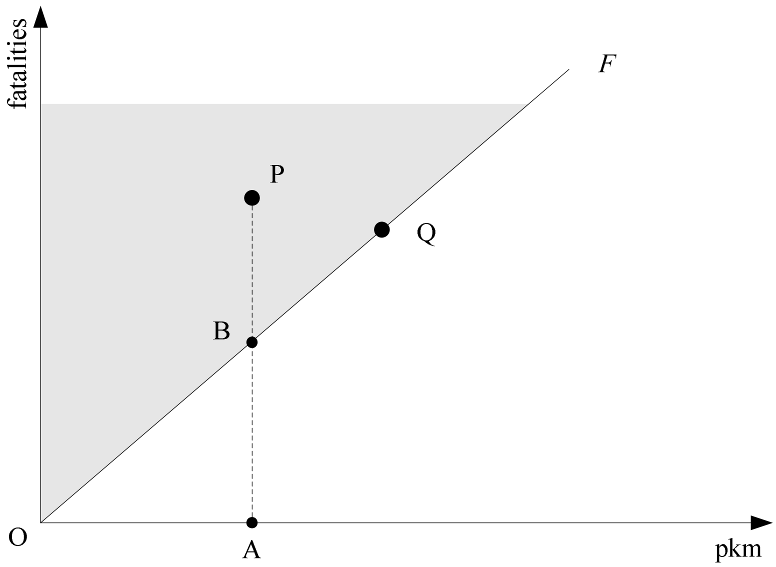


Figure 2.1 Graphic representation of the efficiency production frontier based on the DEA-based road safety model

Mathematically, to use DEA for road safety evaluation, an adjusted road safety output-oriented DEA model[5] is realized as follows:

$$\begin{aligned}
&\max E_0 = \sum_{i=1}^{m} v_i x_{i0} \\
&s.t. \qquad \sum_{r=1}^{s} u_r y_{r0} = 1, \\
&\qquad\qquad \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \le 0, \quad j = 1, \cdots, n \\
&\qquad\qquad u_r, v_i \ge 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m
\end{aligned} \tag{2-1}$$

and

[5] The model can be deduced from the basic input-oriented DEA model (1-4) and (1-5) by switching each of the inputs and outputs into the place of the other.

$$\begin{aligned} \min \quad & \theta_0 \\ s.t. \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \geq x_{io}, \quad i = 1, \cdots, m \\ & \sum_{j=1}^{n} y_{rj}\lambda_j \leq \theta_o y_{ro}, \quad r = 1, \cdots, s \\ & \lambda_j \geq 0, \quad j = 1, \cdots, n \end{aligned} \tag{2-2}$$

where the weighted sum of inputs is to be maximized in the multiplier model (2-1)[6]. The efficiency in DEA is now interpreted as the risk in road safety, and a higher efficiency score means a lower risk. Moreover, solving the corresponding envelopment model (2-2) enables us to find the lowest possible value of θ, for which there exists a HCU that owns at least as much of each input as DMU$_o$, meanwhile leading to no more than θ times each of the outputs of that DMU.

2.4.2 Cross-efficiency model

As indicated in Section 1.5.4, DEA possesses the attractive feature that each DMU is allowed to select its own most favorable input and output weights, or multipliers (v_i^* , u_r^*), for calculating its best efficiency, rather than the same weights for all the DMUs. However, this flexibility in selecting the weights makes the comparison among DMUs on a common base impossible. Moreover, an unreasonable weight scheme could also happen in which some DMUs would heavily weigh a few favorable inputs and outputs and completely ignore others in order to achieve a high relative efficiency score [Dyson & Thannassoulis, 1988; Wong & Beasley, 1990]. To overcome these difficulties, a cross-efficiency method [Sexton et al., 1986] was developed as a DEA extension tool that can be used to identify the best overall performers and to effectively rank all DMUs. Its main idea is to use DEA in a peer evaluation instead of a self-evaluation mode. Specifically, the cross-efficiency method evaluates the performance of a DMU using not only its own optimal input and output weights, but also the ones of all other DMUs. The resulting evaluations can then be aggregated in a cross-

---

[6] In other words, based on the same amount of the road safety final outcomes, we want the exposure to be as high as possible. The objective function can also be the minimization of the weighted sum of outputs. In doing so, however, the efficiency score of each DMU will be equal to or greater than one, rather than lying between zero and one.

efficiency matrix (CEM) as shown in Table 2.3. In the CEM, the element in the *i*th row and *j*th column represents the efficiency score of DMU j using the optimal weights of DMU i. The basic DEA efficiencies are thus located in the leading diagonal. Each column of the CEM is then averaged to obtain a mean cross-efficiency score for each DMU. Since all the DMUs are now evaluated based on the same weighting set, their comparisons can then be made, with a higher cross-efficiency score indicating better overall performance. Moreover, for those DMUs which are probably allocated with unreasonable weights in the basic DEA model, a relatively lower cross-efficiency score will be achieved [Boussofiane et al., 1991]. Therefore, it can also be treated as a kind of sensitivity analysis since different sets of weights are applied to each unit, and they are all internally derived rather than externally imposed. In addition, under some specific conditions, a common set of weights can be deduced from computing the cross-efficiency score [Anderson et al., 2002].

Table 2.3 A generalized cross-efficiency matrix

Rating DMU	Rated DMU				
	1	2	3	...	n
1	E_{11}	E_{12}	E_{13}	...	E_{1n}
2	E_{21}	E_{22}	E_{23}	...	E_{2n}
3	E_{31}	E_{32}	E_{33}	...	E_{3n}
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n	E_{n1}	E_{n2}	E_{n3}	...	E_{nn}
Mean	$\overline{E}_1$	$\overline{E}_2$	$\overline{E}_3$	...	$\overline{E}_n$

One issue that may arise in using the CEM is that the optimal input and output weights obtained from the basic DEA model may not be unique. This makes the cross-efficiency analysis somewhat arbitrary and limits its applicability. During the last decades, some techniques have been proposed for obtaining robust weights for use in the construction of the CEM. The one that is most appropriate for this discussion is known as the aggressive formulation (see Doyle & Green (1994)), which identifies optimal weights that not only maximize the efficiency of the DMU under study but also minimize the sum of efficiencies of all other DMUs. In practice, after calculating the efficiency score of each DMU by using model (2-1), the following formulation is used as the second phase for the cross evaluation of DMU_p:

$$\begin{aligned} \min E_{pj} &= \sum_{i=1}^{m}(v_{ip}\sum_{j\neq p}x_{ij}) \\ s.t. \quad & \sum_{r=1}^{s}(u_{rp}\sum_{j\neq p}y_{rj}) = 1, \\ & \sum_{i=1}^{m}v_{ip}x_{ij} - \sum_{r=1}^{s}u_{rp}y_{rj} \leq 0, \quad \forall j \neq p \\ & \sum_{i=1}^{m}v_{ip}x_{ip} - E_p^*\sum_{r=1}^{s}u_{rp}y_{rp} = 0 \\ & u_r, v_i \geq 0, \quad r = 1,\cdots,s, \quad i = 1,\cdots,m \end{aligned} \tag{2-3}$$

where E_p^* is the efficiency score of DMU_p derived from model (2-1). In model (2-3), the weights will be selected to minimize the efficiency score of a composite DMU (all the DMUs under consideration except DMU_p), given that E_p^* cannot be changed. In this way, the uniqueness of the optimal input and output weights for DMU_p is guaranteed to an utmost extent.

2.4.3 Categorical DEA model

As a remarkable benchmarking approach, DEA owns the capability of indicating a specific reference set for those inefficient DMUs and determining their potential improvement, or target. However, the traditional benchmarking analysis also has certain limitations – an inefficient DMU and its corresponding reference set may not be inherently similar in their practices, or the benchmarks are probably too far away for the inefficient DMU to learn from – which means that the resulted target may not be attainable for this inefficient DMU. To solve this problem, clustering analysis is adopted to first cluster homogeneous DMUs into one group, and the best performer(s) in a particular cluster, derived from a categorical DEA model, is then utilized as benchmark(s) for other DMUs in the same cluster.

Suppose that the n DMUs are clustered into L different categories (K_1, K_2, …, K_L), and the DMUs in K_1 are assumed to be operating under worse circumstances than the ones in K_2, which are worse than the ones in K_3, and so on. The categorical DEA model for road safety evaluation can then be expressed as:

$$
\begin{aligned}
\max E_0 &= \sum_{i=1}^{m} v_i x_{i0} \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{r0} = 1, \\
& \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \le 0, \quad j \in \bigcup_{l=1}^{C} K_l \\
& u_r, v_i \ge 0, \quad r = 1,\cdots,s, \quad i = 1,\cdots,m
\end{aligned}
\tag{2-4}
$$

and

$$
\begin{aligned}
\min \quad & \theta_0 \\
s.t. \quad & \sum_{j \in \bigcup_{l=1}^{C} K_l} x_{ij}\lambda_j \ge x_{io}, \quad i = 1,\cdots,m \\
& \sum_{j \in \bigcup_{l=1}^{C} K_l} y_{rj}\lambda_j \le \theta_o y_{ro}, \quad r = 1,\cdots,s \\
& \lambda_j \ge 0, \quad j = 1,\cdots,n
\end{aligned}
\tag{2-5}
$$

where C denotes the number of categories used in the DEA calculation.

Based on the multiplier form of the categorical model (2-4), we can see that each DMU is now compared only with DMUs in its own and less advantaged categories, rather than with all DMUs as in (2-1). Also, from the envelopment form's point of view, the model (2-5) allows us to evaluate a DMU with respect to the envelopment surface determined for the units contained in it and all preceding categories. Thus, we assess all DMUs $j \in K_1$ with respect to the units in K_1, all DMUs $j \in K_2$ with respect to the units in $K_1 \cup K_2$, and all DMUs $j \in K_C$ with respect to the units in $\bigcup_{l=1}^{C} K_l$.

2.5 Application and Results

In this section, the DEA approach and its extensions are applied to show an overall road safety risk picture of the 29 European countries (the 27 EU Member States together with Switzerland (CH) and Norway (NO)), and to further identify specific benchmarks and to assign practical targets for those underperforming ones in terms of the number of road fatalities. In doing so, the three common measures of exposure to risk, i.e., the number of inhabitants, passenger-kilometres travelled and passenger cars are used as the model's input and the number of road fatalities as output (see Figure 2.2). This structure can be easily extended when other inputs and/or outputs are considered (see also Chapter 4).

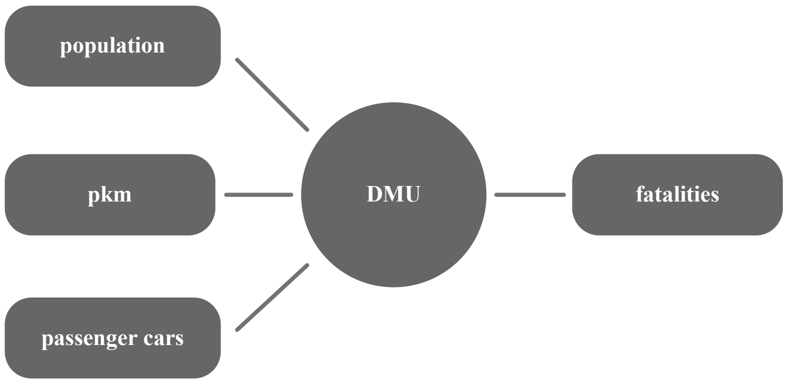


Figure 2.2 The input and output of the model

Data for these 29 European countries are collected from the European Commission (2011a) for the year 2008 and are shown in Table 2.4.

Table 2.4 Data on the three measures of exposure and the road fatalities for the 29 European countries in 2008

Country	Input			Output
	Population (million)	Passenger-kilometres (10 billion)	Passenger cars (million)	Fatalities
BE	10.71	11.04	5.09	944
BG	7.62	4.32	2.22	1061
CZ	10.42	7.24	4.35	1076
DK	5.49	5.22	2.08	406
DE	82.11	87.13	41.25	4477
EE	1.34	1.05	0.54	132
IE	4.43	4.90	1.93	279
EL	11.24	10.00	4.91	1555
ES	45.56	34.26	21.95	3100
FR	62.30	72.02	31.28	4275
IT	59.83	73.68	35.89	4731
CY	0.79	0.58	0.43	82
LV	2.27	1.70	0.92	316
LT	3.36	3.80	1.63	499

LU	0.49	0.67	0.33	35
HU	10.04	4.20	3.03	996
MT	0.41	0.22	0.23	15
NL	16.45	14.70	7.47	677
AT	8.34	7.33	4.27	679
PL	38.13	27.35	15.33	5437
PT	10.62	8.70	4.39	885
RO	21.51	7.05	3.78	3061
SI	2.02	2.49	1.03	214
SK	5.41	2.64	1.49	558
FI	5.31	6.34	2.64	344
SE	9.22	9.84	4.27	397
UK	61.39	67.81	28.96	2645
NO	4.77	5.77	2.18	255
CH	7.65	8.36	3.97	357

Note: Only passenger-kilometres of cars are considered.

Source: European Commission (2011a)

2.5.1 Ranking based on road safety scores

Now, by using the DEA-based road safety model (DEA-RS) as well as the CEM with the aggressive formulation, the overall road safety efficiency score of the 29 European countries can be obtained. They are presented in Table 2.5, together with the standard deviation of each country's 29 efficiency scores. The countries are ranked based on their average cross-efficiency score.

Table 2.5 Overall road safety efficiency score of the 29 European countries based on the DEA-RS model and the CEM

	DEA-RS efficiency	Cross-efficiency	St. dev.
UK	1.000	0.973	0.068
SE	0.991	0.957	0.065
NL	0.988	0.931	0.067
MT	1.000	0.931	0.168
CH	0.966	0.909	0.062
NO	0.883	0.811	0.072
DE	0.802	0.764	0.049
LU	0.806	0.683	0.093
FI	0.719	0.674	0.055
IE	0.683	0.655	0.051

FR	0.663	0.629	0.048
IT	0.652	0.579	0.059
ES	0.574	0.536	0.053
DK	0.559	0.516	0.048
AT	0.500	0.478	0.034
BE	0.479	0.463	0.029
PT	0.478	0.439	0.042
SI	0.454	0.420	0.037
EE	0.400	0.364	0.039
CY	0.381	0.357	0.039
CZ	0.372	0.336	0.043
HU	0.369	0.288	0.082
LT	0.298	0.286	0.021
SK	0.355	0.284	0.077
EL	0.294	0.275	0.021
LV	0.280	0.254	0.029
PL	0.271	0.244	0.031
BG	0.265	0.221	0.050
RO	0.257	0.178	0.075

It can be seen that United Kingdom and Malta are the two best-performing countries since they obtain the optimal efficiency score of one in the DEA-RS model, while the remaining 27 countries (obtaining a value less than one) are considered to be underperforming. Moreover, to represent a true peer assessment for each country, all other countries' best possible weights are utilized to calculate the efficiency score of the country under study, and the average cross-efficiency score is then obtained reflecting this country's all round performance, which is shown in the third column of Table 2.5. Countries can now be ranked by their scores. We can see that the SUN countries (Sweden, United Kingdom, and the Netherlands) are ranked at the top, while most of the Central and Eastern European countries, such as Romania and Bulgaria, are still facing great challenges to improve their road safety performance. Furthermore, comparing the ranking result with the ones in Table 2.1, which are based on the three risk indicators separately, we find that the result from the CEM gives us a global view on the country's road safety performance by taking all three aspects of exposure into account, and yet it is not the simple average of those three rankings. In addition, by computing the standard deviation shown in the last column of Table 2.5, we find that Malta obtains the highest value (0.168), which means that the set of efficiency scores calculated for Malta varies the most from its cross-efficiency score. In other words, Malta has the highest level of uncertainty on its efficiency score, and is probably allocated with unreasonable weights in the DEA-RS model. This result will be verified in the following section.

2.5.2 Country clustering

In the benchmarking analysis, each underperforming country should learn from those best-performing ones so as to improve their road safety performance. However, as can be seen in Table 2.5, the efficiency of most of the Central and Eastern European countries are still far away from that of the SUN countries (all less than half). Therefore, clustering analysis is firstly conducted to group the countries with inherent similarity in their practices. In this study, a hierarchical clustering analysis is applied based on the three exposure measures of each country, which are unitized by the number of road fatalities, respectively. Moreover, three different techniques, i.e., the Ward's method, the Centroid Linkage method, and the Average Linkage method (between and within groups) in SPSS 17.0 are used to derive various clusters. The dendrograms from these three methods are shown in Appendix I.

No matter which clustering method is used, even though certain differences can be found between some countries, five clusters of countries can always be classified, which are listed as follows:

Group 1: BG; CY; CZ; EE; EL; HU; LV; LT; RO; PL; SK
Group 2: AT; BE; DK; PT; SI; ES
Group 3: FI; FR; DE; IE; IT; LU; NO
Group 4: NL; SE; UK; CH
Group 5: MT

They can also be presented in the following 2-dimensional figure indicating the population size and the pkm in each country relative to its number of road fatalities[7].

[7] The inverse risk indicators are applied here because the ratio between exposure and final outcome is maximized in the model.

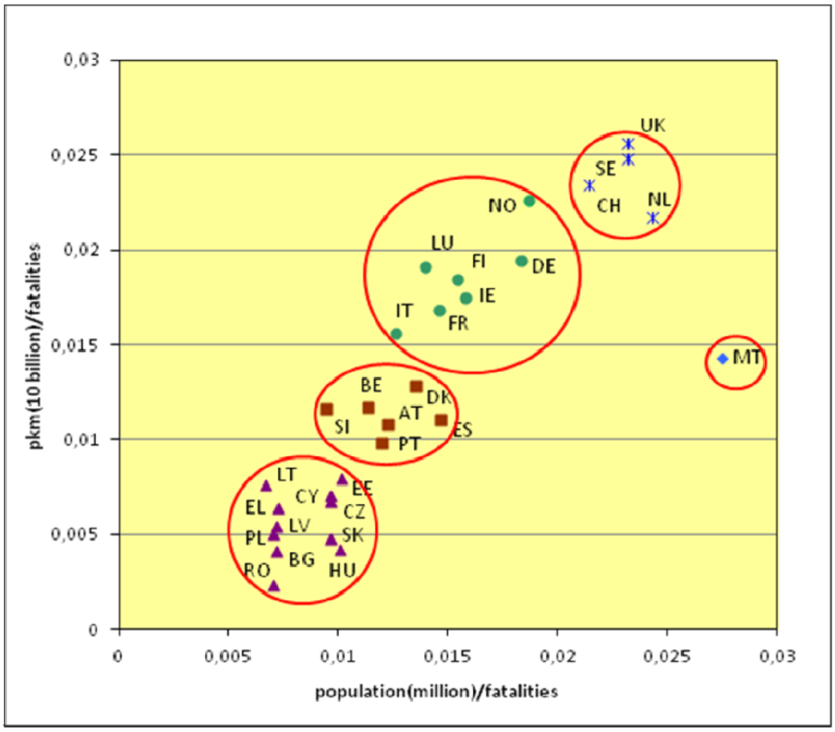


Figure 2.3 29 European countries clustered in five groups

We can see that most of the Central and Eastern European countries are clustered in Group 1, and they have the highest risk on road traffic fatal crashes compared to other countries. The SUN countries and Switzerland are also in one group (Group 4) and they exhibit the best road safety performance. For the remaining countries, three groups are classified, in which Malta appears to be quite different since it is not grouped with any other country. Its distinctive performance also interprets the reason why it obtains the highest standard deviation when calculating the cross-efficiency score (see Table 2.5). Therefore, in the following benchmarking analysis, Malta is excluded and only the four main groups of countries are taken into account.

2.5.3 Benchmarking and target setting

To provide a further evaluation for the 28 European countries, which are clustered in four groups, the categorical DEA-RS model is applied to indicate a specific reference set of benchmarks and to determine the potential improvement for those underperforming countries. More specifically, a country in any group is now compared only with those other countries in the same or less-advantaged group. It means in our case that four DEA-RS models will be created. The first model contains the countries in Group 1 (having the lowest road safety performance), the second model includes the countries belonging to the first two groups, and so on. The results are shown in Table 2.6.

Table 2.6 The efficiency score of the 28 European countries using the categorical DEA-RS model

Country	Category	Efficiency score			
		Category 1	Category 1-2	Category 1-3	Category 1-4
PL	1	0.691	0.477	0.375	0.289
RO	1	0.692	0.478	0.376	0.289
BG	1	0.707	0.489	0.384	0.296
LV	1	0.708	0.489	0.384	0.296
EL	1	0.808	0.531	0.386	0.297
SK	1	0.955	0.660	0.518	0.399
LT	1	0.957	0.604	0.366	0.298
CZ	1	0.959	0.659	0.518	0.399
HU	1	0.993	0.686	0.539	0.415
EE	1	1	0.700	0.543	0.418
CY	1	1	0.741	0.565	0.472
PT	2		0.840	0.642	0.494
SI	2		0.914	0.544	0.454
AT	2		0.944	0.681	0.566
BE	2		0.948	0.611	0.490
DK	2		1	0.723	0.564
ES	2		1	0.794	0.639
IT	3			0.811	0.682
FR	3			0.815	0.667
IE	3			0.846	0.683
FI	3			0.866	0.719
DE	3			1	0.831
LU	3			1	0.848
NO	3			1	0.883
SE	4				0.993
CH	4				1
NL	4				1
UK	4				1

It can be seen that in each DEA-RS model, two or three different efficient countries are identified. For instance, Estonia and Cyprus are the two best-performing countries since they obtain the optimal efficiency score of one in the first model, which means that they are at the top of the countries' performance ranking in Group 1, while the remaining nine countries are considered to be underperforming. Moreover, when the final model is applied in which all the 28 countries are included, United Kingdom, the Netherlands, and Switzerland are the three best performers, while Estonia and Cyprus are then only half on their way to become efficient.

To better understand the computational process leading to the efficiency scores presented in Table 2.6, and especially the reasons why the underperforming countries are unable to obtain a value of one, we further explore the mechanism of the multiplier and envelopment forms of the categorical DEA-RS model, respectively. Theoretically, the multiplier or the primal DEA-RS model is to choose the best possible input and output weights under the imposed restrictions to maximize the efficiency score of a certain country. If the optimal weights of a country *A* under study do not result in a value of one for this country but cause the weighted score of another country *B* in the data set to become one, then the model stops. This implies that country *B* is characterized by a lower risk than country *A* with respect to at least one of the exposure aspects since the efficiency score of *B* is relatively higher with the same set of weights. Therefore, country *A* could take country *B* as an example for improving its road safety performance. From the envelopment or the dual DEA-RS model's point of view, the dual weights, i.e., λ, can be viewed as indicating the amount of technical weight that is attributed by each benchmark country in the construction of an efficient HCU. In other words, the countries with non-zero dual weights make up the reference set for the country under study. Using this principle, the reference sets and dual weights of all the underperforming countries in each group can be identified and the corresponding targets obtained as well. The results are shown in Table 2.7, together with the registered number of fatalities in 2008 and 2009 [European Commission, 2011a].

Table 2.7 Reference sets and targets for the underperforming countries in each group

Country	Category	Reference sets and dual weights				Target fatalities				2008 fatalities	2009 fatalities
		{EE;CY}	{DK; ES}	{DE, LU, NO}	{CH, NL, UK}	T1	T2	T3	T4		
PL	1	{28.455; 0}	{0; 0.837}	{0; 0; 7.994}	{0; 2.318; 0}	3756	2594	2038	1569	5437	4572
RO	1	{16.052; 0}	{0; 0.472}	{0; 0; 4.509}	{0; 1.308; 0}	2119	1464	1150	885	3061	2796
BG	1	{5.687; 0}	{0; 0.167}	{0; 0; 1.597}	{0; 0.463; 0}	751	518	407	314	1061	901
LV	1	{1.666; 0.047}	{0; 0.050}	{0; 0; 0.476}	{0; 0.138; 0}	224	154	121	93	316	254
EL	1	{9.524; 0}	{1.184; 0.112}	{0; 0;2.356}	{0; 0.683; 0}	1257	826	601	463	1555	1453
SK	1	{4.037; 0}	{0; 0.119}	{0; 0; 1.134}	{0; 0.329; 0}	533	368	289	223	558	347
LT	1	{3.619; 0}	{0.636; 0.014}	{0.022; 0.105; 0.309}	{0.018; 0; 0.054}	478	302	183	149	498	370
CZ	1	{6.979; 1.352}	{0; 0.229}	{0; 0; 2.184}	{0; 0.633; 0}	1032	709	557	429	1076	901
HU	1	{7.493; 0}	{0; 0.220}	{0; 0; 2.105}	{0; 0.610; 0}	989	683	537	413	996	822
EE	1	{1; 0}	{0.039; 0.025}	{0; 0; 0.281}	{0; 0.081; 0}		92	72	55	132	100
CY	1	{0; 1}	{0; 0.020}	{0.007; 0.395; 0}	{0.108; 0; 0}		61	46	39	82	71
PT	2		{0.654; 0.154}	{0; 0; 2.226}	{0; 0.646; 0}		744	568	437	885	840
SI	2		{0.447; 0,005}	{0.002; 0.933; 0.291}	{0; 0; 0.037}		196	116	97	214	171
AT	2		{0.337; 0.163}	{0.096; 0.957; 0}	{0.973; 0.054; 0}		641	463	384	679	633
BE	2		{1.569; 0.083}	{0.052; 0; 1.343}	{0.281; 0.220; 0.080}		895	577	462	944	955
DK	2		{1; 0}	{0; 0; 1.151}	{0; 0.243; 0.024}			293	229	406	303
ES	2		{0; 1}	{0.303; 0; 4.337}	{2.542; 1.588; 0}			2462	1982	3100	2605
IT	3			{0.313; 69.575; 0}	{9.040; 0; 0}			3838	3227	4731	4050
FR	3			{0.361; 13.784; 5.425}	{1.306; 0; 0.901}			3483	2850	4275	4273
IE	3			{0; 0; 0.929}	{0; 0; 0.072}			237	191	279	240
FI	3			{0.016; 1.448; 0.689}	{0; 0; 0.093}			298	247	344	279
DE	3			{1; 0; 0}	{7.631; 0.830; 0.164}				3720	4477	4152
LU	3			{0; 1; 0}	{0.083; 0; 0}				30	35	47
NO	3			{0; 0; 1}	{0; 0; 0.085}				225	255	212
SE	4				{0; 0.099; 0.124}				394	397	355
CH	4				{1; 0; 0}					357	349
NL	4				{0; 1; 0}					677	644
UK	4				{0; 0; 1}					2645	2337

Firstly, it can be seen that the reference sets for each underperforming country are solely comprised of the best-performing countries in each model. Taking the first one as an example, Estonia acts as a benchmark country for all the nine other countries within this group while Cyprus for two of them.

Moreover, since the value of the dual weight points out the extent to which each benchmark country contributes to the definition of the HCU for each underperforming country, it enables us to rank the benchmark countries in terms of their relative importance. Taking Latvia as an example, the dual weight of Estonia (1.666) is much larger than that of Cyprus (0.047) implying that Estonia plays a stronger role in determining the ideal performance of Latvia.

More importantly, the constructed HCU offers information for setting a practical target for each underperforming country in order to become efficient. In other words, for each underperforming country, a quantitative road safety target can be formulated by learning from its benchmarks, using the following formula:

$$T_j = \sum_{k=1}^{K} \lambda_{Bk} F_{Bk} \quad j = 1, \cdots n \tag{2-6}$$

where T_j denotes the target number of fatalities for the jth underperforming country, K denotes the number of benchmarks in the referent set, F_{Bk} is the number of road fatalities in the kth benchmark country, and λ_{Bk} is the corresponding dual weight. Thus, for the case of Latvia, its target number of fatalities can be calculated as: $1.666 \times 132 + 0.047 \times 82 = 224$. Moreover, based on the different benchmark countries identified in each model, which are Denmark and Spain in the second model, Germany, Luxembourg and Norway in the third model, and Switzerland, United Kingdom and the Netherlands in the final model, three additional targets for the number of fatalities can be obtained for Latvia, which are 154, 121, and 93, respectively. We can see that the first target using Estonia and Cyprus as reference set is more realistic since these two countries are more inherently similar to Latvia in their practices and the assigned target value is more close to its current number of fatalities. However, all other target fatalities are also valuable for Latvia to set its mid- and long-term targets.

To further illustrate the effectiveness of the assigned target(s) in Table 2.7, we compare the results with the actual target of a particular country in Table 2.2. Taking Belgium as an example, which has set a real target of maximum 500 fatalities by 2015, we find that the number is located within the range of targets calculated in this study (i.e., 885, 584, and 461), and actually it is quite close to the last value which is derived by using the overall best-performing countries, i.e., United Kingdom and the Netherlands, as its reference set.

Finally, by checking the number of fatalities of these 28 European countries in 2009 (see the last column of Table 2.7), we find that 9 countries (Austria, Czech Republic, Finland, Hungary, Lithuania, Norway, Slovakia, Slovenia, and Sweden) have already achieved their first target, and Slovakia has even reached its second target. It means that the target setting approach used is practical and attainable. However, it should also be mentioned here that every year, the performance of each country changes and the benchmark countries could alter as well. Thus, the target value for each country should also modify correspondingly. Therefore, it is a dynamic process with respect to both target setting and target achieving.

2.6 Conclusion

In this chapter, the road safety performance of 29 European countries has been evaluated by considering three main risk indicators (i.e., the number of fatalities per million inhabitants, the number of fatalities per 10 billion passenger-kilometres travelled, and the number of fatalities per million passenger cars) simultaneously, and useful benchmarks are identified for improving the operations of those underperforming countries by assigning practical targets for them. In doing so, data envelopment analysis and its three model extensions were investigated. First, a particular DEA-based road safety model was proposed to solve the problem that the outputs such as the number of fatalities need to be minimized rather than maximized in road safety risk evaluation. This can be treated as a natural extension of the basic DEA model. Second, the cross-efficiency method was used to enable the comparison among countries on a common basis. The best overall performers could then be identified and all countries ranked. Third, the clustering analysis was conducted to group countries with inherent similarity in their practices, and the categorical DEA-RS model was applied to indicate appropriate benchmarks within each cluster and determine the potential degree of improvement for those underperforming countries.

The analysis, which was based on the combination of these model extensions, provided interesting insights and valuable recommendations for road safety policymakers in identifying the efficiency of their current operations and in suggesting targets for improvement. Specifically, using the DEA-RS model linking input (three measures of exposure to risk) and output (the number of fatalities), an overall road safety efficiency score was obtained for each country and the ranking of these countries was deduced by computing their cross-efficiency. We found that United Kingdom, Sweden, and the Netherlands were three 'SUN' countries possessing the best road safety performance among all the European countries in 2008, while most of the Central and Eastern European countries were still facing great challenges in this aspect. Moreover, based on the results from the hierarchical clustering analysis, four groups of these European countries (except Malta) were classified with their road safety performance from worst to best. Then, by applying the categorical DEA-RS model, best-performing and underperforming countries were identified in each group, and reference sets or benchmarks for those underperforming ones were indicated. More importantly, the extent to which each reference set could be learned from was specified, and practical targets on fatalities were given for each underperforming country. They enable policymakers to recognize the gap with those best-performing countries and further develop their own road safety policy. In this study, we found that United Kingdom, the Netherlands, and Switzerland were the three benchmark countries, and could be used to derive potential improvement for all others. However, from the practical point of view, for those countries with a relatively high fatal risk, such as Latvia, the reference set consisting of for example Estonia and Cyprus was more suitable to be learned from, at least in the short term. Thus the target number of fatalities deduced from this reference set was more realistic and attainable. Nevertheless, other challenging targets are also valuable, especially for the long-term development.

To conclude, it would be interesting to perform in the future an empirical investigation on whether underperforming countries would choose the specific benchmarks indicated in this study as it will help in determining the validity of the methodology. Also, from the road safety policy point of view, we should keep in mind that setting targets does not guarantee their achievement unless keeping adequate political ambition, effective strategies, sufficient allocation of resources, successful implementation, and persistent monitoring and evaluation as an ongoing process throughout the whole target period.

Chapter 3 Road Safety Development in Europe: A Decade of Changes (2000-2009) [8]

This chapter answers the third research question of this dissertation, i.e., how to evaluate the road safety performance change of countries over time. In doing so, we not only focus on the evolution in the number of road fatalities within a given period, but also take the change in exposure in the same period into account. The DEA-based road safety model and the Malmquist productivity index are employed to undertake the assessment.

3.1 Introduction

In the previous chapter, *cross-sectional data* analysis based on DEA and its several extensions was conducted to evaluate the road safety performance of 28 European countries at one specific point of time, which was the year 2008. As a result, countries were ranked and benchmarked based on their relative road safety performance in that year. In other words, data with respect to only one time period were analyzed, while the evolution of road safety performance in each country was out of consideration. In doing so, it is not possible to reflect in the results such as that the best-performing countries in one year might not be the best in the previous or next year. Consequently, research on *time series cross-sectional data* collected at regular intervals, also known as *longitudinal data* analysis, is required to gain a clear understanding of the road safety development in each country relative to others. Interesting questions can then be asked that can never be answered with pure cross-sectional data, such as the performance change of countries over time.

[8] Related research has been published in: Shen, Y., Ruan, D., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K., (2011). Sustainable road transport in the European Union: Changes in undesirable impacts, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2242, pp. 37-44.

Traditionally, the percentage change in the number of people killed on the road is the main indicator used to compare the development of road safety between countries with a higher reduction in road fatalities indicating a better rank [European Transport Safety Council, 2011] (see Figure 3.1).

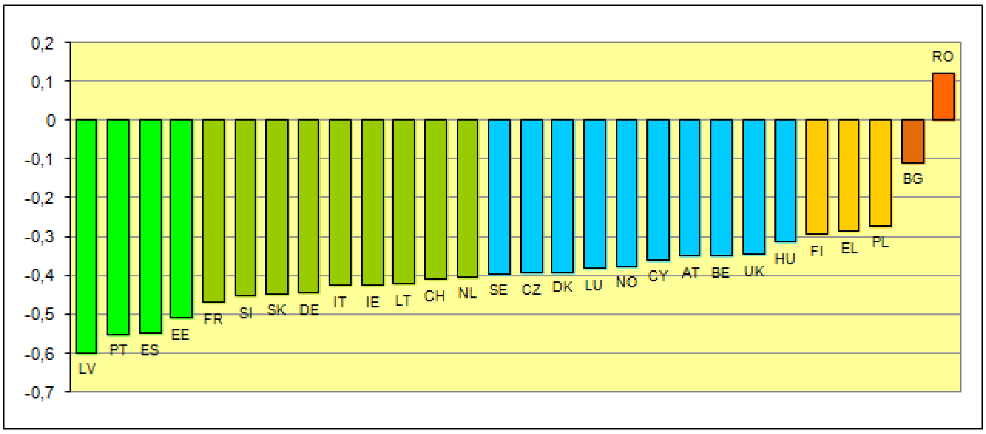


Figure 3.1 Fatality change of 28 European countries between 2000 and 2009

Data source: European Commission (2011a)

The idea is intuitive and the results are easy to obtain since the number of road fatalities in two years is the only information needed for the calculation. However, simply considering the reduction in the final outcome may not correctly reflect the real improvement in road safety because the transport circumstances of a country underlying the final outcome also change every year. For instance, consider a country that recorded 100 road fatalities in one year with a participation of 10 billion pkm in traffic, and 90 road fatalities with 9 billion pkm in the second year. Although the number of fatalities is reduced by 10% between these two years, there is actually no improvement in road safety performance because the degree of participation in traffic also decreases by 10% in this country and its fatality risk has thereby not changed during these two years. Consequently, to capture the dynamic road safety progress in each country, this study not only focuses on the development of road fatalities within a given period, but also takes the change in exposure into account for the three measures used in the previous chapter, which are the number of inhabitants, passenger-kilometres travelled and passenger cars. The DEA-based road safety model proposed in Section 2.4.1 and the Malmquist productivity index

[Malmquist, 1953] are employed to undertake the assessment for the same 28 European countries over the period 2000-2009.

The remaining of this chapter is organized as follows. In Section 3.2, we elaborate the construction of the Malmquist productivity index based on the DEA-based road safety model. In Section 3.3, we demonstrate the application of this DEA-based Malmquist productivity index for measuring the road safety development of countries over time, and the results are subsequently provided and discussed. The chapter is summarized in Section 3.4.

3.2 DEA-based Malmquist Productivity Index

The concept of the Malmquist productivity index, originally introduced by Malmquist (1953) as a quantity for analyzing the consumption of inputs, has been further developed by Caves et al. (1982). Afterwards, Färe et al. (1992) combined the ideas on the measurement of efficiency and the measurement of productivity to construct a Malmquist productivity index directly from input and output data using DEA. Specifically, by using longitudinal data, the *DEA-based Malmquist productivity index*, hereafter referred to as DEA-MI, relies on firstly constructing an efficiency production frontier over the whole sample realized by DEA (as illustrated in Section 1.5.2), and then computing the distance of individual observations to the frontier. In practice, the DEA-MI has proven to be a proper tool for measuring the productivity change of DMUs over time [Chen & Ali, 2004; Yörük & Zaim, 2005; Greer, 2008].

Moreover, in contrast to conventional production functions or other index approaches, the DEA-MI can be further decomposed into two components, one measuring the *change in efficiency* (*EFFCH*) and the other indicating the *change in the frontier technology* (*TECHCH*). From the output-oriented view of road safety development assessed in this study, an improvement in efficiency occurs when there are decreases in the quantities of output (i.e., road fatalities) based on a given set of inputs. Operationally, it can for example be realized by more and better road user education and driver training. Moreover, encouraging citizens to use public transport instead of private cars is also widely recognized as a useful way in reducing the road crash risk in a country. In contrast to a change in efficiency, technical change occurs through the adoption of new technologies that reduce the *minimum* quantities of output given a certain level of input. In this respect, the introduction of safer vehicles, betterment of road infrastructure, and improvement in medical treatment of people involved in crashes are all related to productivity-enhancing technical changes.

Towards a safer use of the road, both efficiency enhancements and technical improvements are required. The DEA-MI calculated here allows us to measure the combined effect of *EFFCH* and *TECHCH* of each country within the given period, and it also captures the separate impact of each effect.

Mathematically, the DEA-MI is computed as the product of *EFFCH* and *TECHCH*. Therefore, to obtain the total factor productivity change of a DMU over time, we need to firstly derive its *EFFCH* and *TECHCH*, respectively. In doing so, consider the same example as presented in Figure 2.1, but with two time periods t and t+1 for the two units P(x_0, y_0) and Q(x_1, y_1) now, which is illustrated in Figure 3.2.

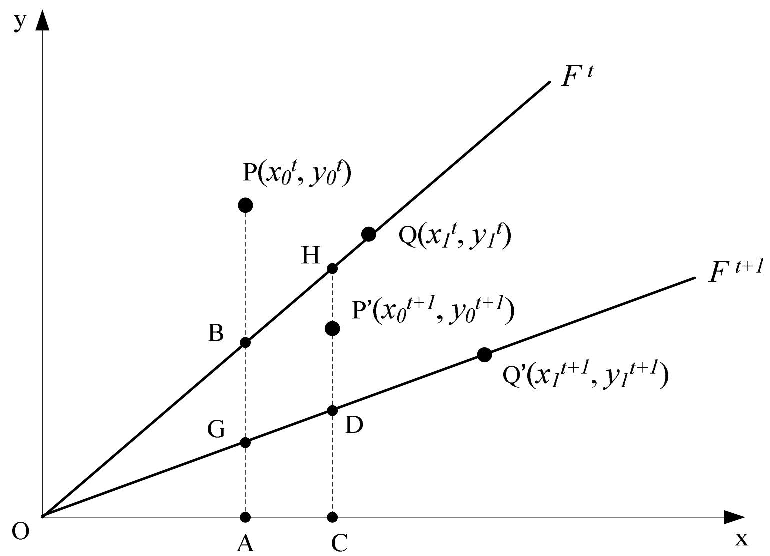


Figure 3.2 Graphic representation for *EFFCH* and *TECHCH* computation

By identifying the efficient unit in each time period, which is Q(x_1^t, y_1^t) and Q'(x_1^{t+1}, y_1^{t+1}), respectively, we derive the efficiency frontiers F^t and F^{t+1} as in Figure 3.2, and the efficiency score of unit P in each time period can be measured as AB/AP and CD/CP', respectively. Thus, the magnitude of the efficiency change of unit P from period *t* to *t*+1 can be computed as the ratio of these two efficiency scores, which can be further expressed in the corresponding *distance function* forms as follows:

$$EFFCH = \frac{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{D_o^t(x_o^t, y_o^t)} \tag{3-1}$$

where the two distance functions can be computed by means of the DEA-based road safety model as in (2-2), and they are represented as below:

$$\begin{aligned} & D_o^t(x_o^t, y_o^t) = \min \theta \\ & s.t. \quad \sum_{j=1}^{n} x_{ij}^t \lambda_j \geq x_{io}^t, \quad i = 1, \cdots, m \\ & \qquad \sum_{j=1}^{n} y_{rj}^t \lambda_j \leq \theta y_{ro}^t, \quad r = 1, \cdots, s \\ & \qquad \lambda_j \geq 0, \quad j = 1, \cdots, n \end{aligned} \tag{3-2}$$

and

$$\begin{aligned} & D_o^{t+1}(x_o^{t+1}, y_o^{t+1}) = \min \theta \\ & s.t. \quad \sum_{j=1}^{n} x_{ij}^{t+1} \lambda_j \geq x_{io}^{t+1}, \quad i = 1, \cdots, m \\ & \qquad \sum_{j=1}^{n} y_{rj}^{t+1} \lambda_j \leq \theta y_{ro}^{t+1}, \quad r = 1, \cdots, s \\ & \qquad \lambda_j \geq 0, \quad j = 1, \cdots, n \end{aligned} \tag{3-3}$$

where x_{ij}^t, y_{rj}^t, x_{ij}^{t+1}, and y_{rj}^{t+1} denote the *i*th input and *r*th output of the *j*th DMU at a given point in time *t* and *t*+1, respectively.

For the effect of the efficiency change, which also reflects the capability of an inefficient DMU in catching up with those efficient ones, *EFFCH*>1 indicates progress in the relative efficiency of the DMU_o from period *t* to *t*+1, while *EFFCH*=1 and *EFFCH*<1 means respectively no change and regress in efficiency.

To fully evaluate the total factor productivity change, we should also take into account the technical change, which measures the shift in the technology frontier between two time periods. In Figure 3.2, we notice that the production possibilities set expands from period *t* to *t*+1, as a greater number of input-output combinations become feasible when the frontier moves from F^t to F^{t+1}, and the HCU of unit P also moves from B to G. Thus, the *TECHCH* at $\mathrm{P}(x_0^t,\ y_0^t)$ is evaluated by: AB/AG, which is equivalent to:

$$TECHCH_{\mathrm{P}} = \frac{\mathrm{AB}/\mathrm{AP}}{\mathrm{AG}/\mathrm{AP}} = \frac{D_o^t(x_o^t, y_o^t)}{D_o^{t+1}(x_o^t, y_o^t)} \tag{3-4}$$

where the denominator $D_o^{t+1}(x_o^t, y_o^t)$ denotes the relative efficiency of $\mathrm{P}(x_0^t,\ y_0^t)$ with respect to the frontier at time t+1, i.e., F^{t+1}.

Similarly, the *TECHCH* at $\mathrm{P'}(x_0^{t+1},\ y_0^{t+1})$ can be expressed by:

$$TECHCH_{\mathrm{P'}} = \frac{\mathrm{CH}/\mathrm{CP'}}{\mathrm{CD}/\mathrm{CP'}} = \frac{D_o^t(x_o^{t+1}, y_o^{t+1})}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \tag{3-5}$$

where the numerator $D_o^t(x_o^{t+1}, y_o^{t+1})$ represents the relative efficiency of $\mathrm{P'}(x_0^{t+1},\ y_0^{t+1})$ relative to the frontier at time t, i.e., F^t.

The overall *TECHCH* is therefore defined as the geometric mean[9] of the above two *TECHCH*s:

$$TECHCH = \left[\frac{D_o^t(x_o^t, y_o^t)}{D_o^{t+1}(x_o^t, y_o^t)} \frac{D_o^t(x_o^{t+1}, y_o^{t+1})}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \right]^{1/2} \tag{3-6}$$

where the two mixed-period measures, i.e., $D_o^{t+1}(x_o^t, y_o^t)$ and $D_o^t(x_o^{t+1}, y_o^{t+1})$, can be derived from the following modification of the DEA-based road safety model as in (2-2):

$$\begin{aligned} & D_o^{t+1}(x_o^t, y_o^t) = \min \theta \\ & s.t. \quad \sum_{j=1}^{n} x_{ij}^{t+1} \lambda_j \geq x_{io}^t, \quad i = 1, \cdots, m \\ & \qquad\quad \sum_{j=1}^{n} y_{rj}^{t+1} \lambda_j \leq \theta y_{ro}^t, \quad r = 1, \cdots, s \\ & \qquad\quad \lambda_j \geq 0, \quad j = 1, \cdots, n \end{aligned} \tag{3-7}$$

and

[9] It has to be noted that the geometric mean version of the Malmquist productivity index does not satisfy the circular test [Pastor & Lovell, 2007].

$$
\begin{aligned}
& D_o^t(x_o^{t+1}, y_o^{t+1}) = \min \theta \\
& s.t. \quad \sum_{j=1}^{n} x_{ij}^t \lambda_j \geq x_{io}^{t+1}, \quad i = 1, \cdots, m \\
& \qquad \sum_{j=1}^{n} y_{rj}^t \lambda_j \leq \theta y_{ro}^{t+1}, \quad r = 1, \cdots, s \\
& \qquad \lambda_j \geq 0, \quad j = 1, \cdots, n
\end{aligned}
\tag{3-8}
$$

For the change in the frontier technology, values greater than one indicate an improvement in this aspect, while values equal to and less than one imply status quo and deterioration, respectively.

By now, the DEA-MI, which measures the total factor productivity change of a particular DMU$_o$ from period t to period t+1, can be computed as the product of *EFFCH* and *TECHCH*:

$$
\begin{aligned}
MI_o &= \frac{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{D_o^t(x_o^t, y_o^t)} \left[\frac{D_o^t(x_o^t, y_o^t)}{D_o^{t+1}(x_o^t, y_o^t)} \frac{D_o^t(x_o^{t+1}, y_o^{t+1})}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \right]^{1/2} \\
&= \left[\frac{D_o^t(x_o^{t+1}, y_o^{t+1})}{D_o^t(x_o^t, y_o^t)} \frac{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{D_o^{t+1}(x_o^t, y_o^t)} \right]^{1/2}
\end{aligned}
\tag{3-9}
$$

MI_o>1 indicates progress in the total factor productivity of the DMU$_o$ from period t to t+1, while MI_o=1 and MI_o<1 means respectively status quo and decay in productivity.

In the following section, the Malmquist productivity index based on the DEA-based road safety model (DEA-RS-MI) is applied to evaluate the road safety performance change of countries over time. Meanwhile, the two effects on efficiency enhancements and technical improvements are measured separately for country comparisons.

3.3 Application and Results

In this study, the evolution in the number of road fatalities and the changes in three common measures of exposure to risk used in the previous chapter, i.e., the number of inhabitants, passenger-kilometres travelled and passenger cars, are considered simultaneously in order to assess the dynamic road safety progress in Europe during the last decade. The input-output structure is the same as in Figure 2.2, in which the three measures of exposure are used as the model's input and the number of road fatalities as output. Data are collected from 2000 to 2009 (the latest year for which data are available) for the same 28 European countries as considered in Chapter 2 [European Commission, 2011a]. The data for 2000 and 2009 are shown in Table 3.1.

Table 3.1 Input and output data of the 28 European countries for 2000 and 2009

Country	Input						Output	
	Population (million)		Passenger-kilometres (10 billion)		Passenger cars (million)		Fatalities	
	2000	2009	2000	2009	2000	2009	2000	2009
BE	10.25	10.80	10.55	11.15	4.63	5.16	1470	955
BG	8.17	7.59	2.69	4.63	1.95	2.43	1012	901
CZ	10.27	10.49	6.39	7.23	3.44	4.43	1486	901
DK	5.34	5.52	5.06	5.22	1.85	2.11	498	303
DE	82.21	81.90	83.13	88.68	38.74	41.53	7503	4152
EE	1.37	1.34	0.67	1.05	0.46	0.55	204	98
IE	3.81	4.46	3.84	4.83	1.31	1.94	418	240
EL	10.92	11.28	6.30	10.13	3.06	5.08	2037	1453
ES	40.26	45.91	30.26	35.05	17.15	22.06	5777	2714
FR	59.06	62.63	69.96	72.39	29.54	31.25	8079	4273
IT	56.94	60.19	72.65	70.81	32.31	36.29	7061	4237
CY	0.69	0.80	0.39	0.60	0.26	0.45	111	71
LV	2.37	2.25	1.15	1.67	0.54	0.92	635	254
LT	3.50	3.34	2.60	3.61	1.13	1.68	641	370
LU	0.44	0.50	0.56	0.67	0.27	0.33	76	47
HU	10.21	10.02	4.62	4.12	2.31	3.03	1200	822
NL	15.93	16.53	14.11	14.63	6.44	7.58	1082	644
AT	8.01	8.37	6.67	7.23	4.05	4.32	976	633
PL	38.45	38.15	14.97	28.50	9.64	16.29	6294	4572
PT	10.23	10.63	7.10	8.60	3.40	4.43	1877	840
RO	22.44	21.48	5.10	7.55	2.74	4.14	2466	2796
SI	1.99	2.04	2.03	2.49	0.86	1.05	314	171
SK	5.39	5.42	2.39	2.64	1.26	1.57	628	384
FI	5.18	5.34	5.57	6.43	2.11	2.74	396	279
SE	8.87	9.30	9.19	9.94	3.94	4.29	591	358
UK	58.89	61.80	63.97	68.02	24.85	29.10	3580	2337
NO	4.49	4.83	5.12	5.83	1.83	2.22	341	212
CH	7.18	7.74	7.86	8.49	3.51	4.00	592	349

Source: European Commission (2011a)

3.3.1 The overall results

The DEA-RS-MI is now adopted to measure the extent to which the 28 European countries have improved their level of road safety performance during the period under study. The overall results are shown in Figure 3.3.

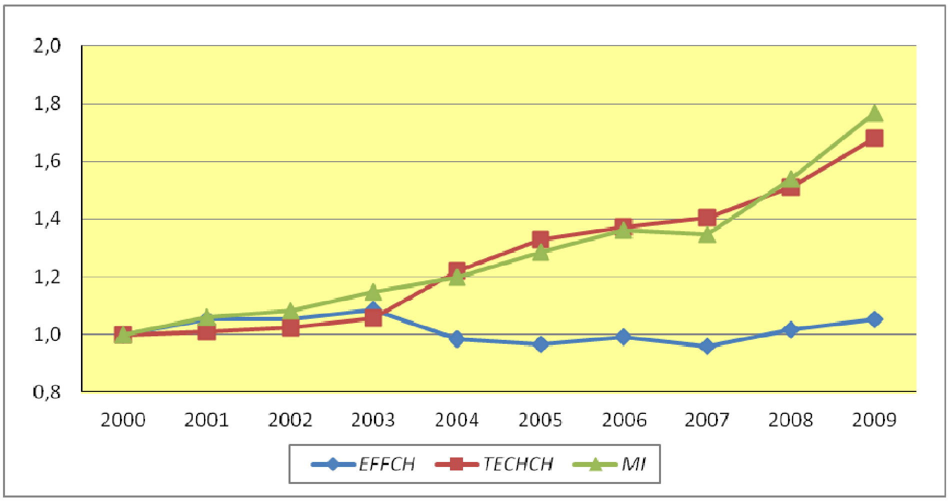


Figure 3.3 The evolution in *MI* of the 28 European countries and its decomposition into technical and efficiency changes in 2000-2009

Figure 3.3 indicates the cumulative *MI* of the 28 European countries and its decomposition (i.e., in *EFFCH* and *TECHCH*) from 2000 to 2009 by sequential multiplication of the improvements in each year with 2000 as the index year (equal to one). From the trend of *MI*, we can see that the 28 European countries as a whole exhibit considerable improvement in road safety performance (nearly 80%) during the last decade. Although a slight decrease existed in 2007, the total 'productivity' went steadily up during this period, and the trend was much steeper in the last two years. Moreover, we can find that this was mostly dominated by its technical component, which means that the main source of this growth came about more through the adoption of productivity-enhancing new technologies throughout the road transport sector in Europe than through the efficiency improvements among those relatively inefficient countries. More specifically, based on the trends of *EFFCH* and *TECHCH*, we can see that both efficiency and technology in these 28 European countries were improved during the first three years of the last decade (2001-2003), and those underperforming countries did even a little better in catching up with those efficient ones as the *EFFCH* is somewhat greater than the *TECHCH*. However, in the following four years (2004-2007), great efforts have been made in Europe to update its road safety technology, resulting in a remarkable shift in the technology frontier. On the other hand, the countries seemed to have lost their momentum by 2007 for further improvement on their efficiency and those underperforming countries showed difficulties in keeping pace with their benchmarks. Finally, in 2008 and 2009, both underperforming and best-performing countries made progress in terms of their road safety efficiency and technology together again, which enabled the total factor productivity to increase dramatically.

3.3.2 Cross-country comparisons

Although considerable improvement in terms of road safety performance has been achieved in Europe during the last decade, the situation differs widely from country to country. Therefore, apart from analyzing the road safety development of the 28 European countries by considering them as a whole, the progress in each of these countries is illustrated in Appendix II, and the cross-country comparisons are provided in the following sections.

3.3.2.1 Efficiency change

To compare the road safety progress in these 28 European countries during the last decade, we firstly look at the changes in their relative efficiency. Tables 3.2 and 3.3 present the DEA-RS efficiency scores and the corresponding efficiency changes of the 28 European countries over the period 2000-2009. It can be seen from Table 3.2 that the Netherlands, Sweden, United Kingdom, Norway, and Switzerland were the five best-performing countries in terms of road safety since they obtained the efficiency score of one alternatively during these ten years. In other words, they determined the efficiency levels of other countries since they were the ones that shifted the frontier in this period. The remaining countries, however, had an efficiency score less than one in each time period, and both improvement and decline occurred during these ten years. Within these countries, there were still seven, i.e., Belgium, Bulgaria, Greece, Hungary, Austria, Poland, and Romania, whose overall efficiency (2009 compared to 2000) changed less than one (see the last column of Table 3.3), which implies their weak capability in catching up with those efficient countries. On the contrary, comparison of development up to 2009 shows that Spain, Latvia, and Portugal achieved the best improvement in terms of efficiency (all over 40%), which could be mainly attributable to the prominent reduction in their road fatalities during this period (see Figure 3.1). In addition, it should be noted that their poor efficiency scores in 2000 (0.43, 0.23, and 0.33, respectively) also provided them with more space for progress.

Table 3.2 Efficiency scores of the 28 European countries over the period 2000-2009

Country	Efficiency score									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
BE	0.45	0.45	0.50	0.52	0.49	0.48	0.48	0.47	0.49	0.43
BG	0.49	0.48	0.50	0.48	0.41	0.37	0.33	0.33	0.30	0.32
CZ	0.42	0.47	0.43	0.42	0.37	0.37	0.43	0.37	0.40	0.44
DK	0.65	0.76	0.70	0.74	0.73	0.76	0.80	0.59	0.56	0.69
DE	0.74	0.80	0.79	0.79	0.79	0.80	0.78	0.81	0.83	0.80
EE	0.41	0.42	0.37	0.49	0.39	0.37	0.29	0.30	0.42	0.52
IE	0.55	0.57	0.63	0.70	0.57	0.50	0.57	0.61	0.68	0.70
EL	0.33	0.36	0.41	0.41	0.33	0.31	0.30	0.30	0.30	0.29
ES	0.43	0.46	0.47	0.46	0.47	0.48	0.50	0.54	0.64	0.65
FR	0.53	0.52	0.55	0.67	0.66	0.61	0.69	0.69	0.67	0.59
IT	0.66	0.66	0.66	0.68	0.65	0.63	0.59	0.69	0.68	0.69
CY	0.38	0.44	0.46	0.44	0.32	0.36	0.42	0.43	0.47	0.51
LV	0.23	0.26	0.25	0.26	0.22	0.24	0.25	0.24	0.30	0.34
LT	0.33	0.30	0.30	0.29	0.23	0.21	0.23	0.23	0.30	0.37
LU	0.51	0.56	0.63	0.71	0.70	0.70	0.82	0.67	0.84	0.56
HU	0.52	0.50	0.43	0.45	0.39	0.36	0.35	0.35	0.42	0.46
NL	0.90	0.98	0.99	0.93	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.97
AT	0.60	0.61	0.59	0.56	0.54	0.57	0.55	0.59	0.57	0.55
PL	0.37	0.42	0.40	0.40	0.33	0.32	0.33	0.30	0.29	0.32
PT	0.33	0.38	0.38	0.40	0.40	0.39	0.49	0.47	0.49	0.48
RO	0.55	0.55	0.55	0.57	0.44	0.40	0.39	0.33	0.29	0.29
SI	0.39	0.45	0.45	0.49	0.40	0.39	0.39	0.36	0.45	0.50
SK	0.52	0.53	0.53	0.49	0.44	0.44	0.42	0.37	0.40	0.53
FI	0.80	0.73	0.76	0.84	0.78	0.69	0.82	0.71	0.72	0.79
SE	0.96	0.97	0.99	**1.00**	**1.00**	**1.00**	**1.00**	0.93	0.99	0.98
UK	**1.00**	**1.00**	**1.00**	0.99	0.99	0.90	0.93	0.97	**1.00**	**1.00**
NO	0.84	**1.00**	0.91	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	0.88	0.95
CH	0.85	0.93	0.98	0.89	0.86	**1.00**	**1.00**	**1.00**	**1.00**	0.92

Table 3.3 Efficiency changes of the 28 European countries from 2000 to 2009

Country	*EFFCH*									
	01/00	02/01	03/02	04/03	05/04	06/05	07/06	08/07	09/08	09/00
BE	0.99	1.11	1.03	0.94	0.99	0.99	0.99	1.04	0.89	0.96
BG	0.98	1.02	0.97	0.85	0.91	0.89	1.00	0.90	1.07	0.65
CZ	1.11	0.92	0.97	0.88	1.00	1.18	0.85	1.09	1.10	1.05
DK	1.16	0.93	1.05	0.98	1.04	1.06	0.74	0.95	1.22	1.06
DE	1.07	1.00	0.99	1.01	1.01	0.97	1.04	1.02	0.97	1.08
EE	1.02	0.88	1.32	0.80	0.93	0.80	1.01	1.41	1.24	1.27
IE	1.03	1.11	1.11	0.81	0.87	1.14	1.07	1.13	1.03	1.27
EL	1.09	1.14	1.00	0.81	0.94	0.97	1.00	0.99	0.99	0.90
ES	1.07	1.04	0.97	1.02	1.02	1.04	1.09	1.18	1.02	1.53
FR	0.99	1.05	1.22	0.99	0.93	1.12	1.01	0.96	0.88	1.11
IT	1.00	1.00	1.03	0.96	0.97	0.94	1.16	0.99	1.01	1.04
CY	1.15	1.05	0.96	0.72	1.14	1.16	1.02	1.10	1.09	1.34
LV	1.13	0.98	1.02	0.86	1.08	1.05	0.94	1.26	1.14	1.48
LT	0.90	1.00	0.96	0.78	0.92	1.11	0.99	1.31	1.22	1.10
LU	1.10	1.12	1.14	0.98	1.00	1.17	0.82	1.24	0.68	1.11
HU	0.97	0.86	1.05	0.85	0.94	0.95	1.02	1.18	1.11	0.89
NL	1.10	1.00	0.94	1.07	1.00	1.00	1.00	1.00	0.97	1.08
AT	1.02	0.96	0.96	0.96	1.07	0.96	1.08	0.95	0.97	0.92
PL	1.13	0.94	1.01	0.83	0.98	1.01	0.91	0.98	1.09	0.85
PT	1.13	1.01	1.06	1.00	0.97	1.25	0.97	1.05	0.97	1.45
RO	0.99	1.00	1.05	0.77	0.91	0.96	0.86	0.87	1.01	0.53
SI	1.13	1.02	1.08	0.82	0.99	1.00	0.91	1.27	1.10	1.27
SK	1.02	1.00	0.93	0.89	1.00	0.94	0.90	1.07	1.34	1.02
FI	0.92	1.04	1.11	0.94	0.89	1.19	0.86	1.01	1.10	1.00
SE	1.01	1.01	1.02	1.00	1.00	1.00	0.93	1.07	0.99	1.02
UK	1.00	1.00	0.99	0.99	0.91	1.03	1.05	1.03	1.00	1.00
NO	1.19	0.91	1.10	1.00	1.00	1.00	1.00	0.88	1.07	1.13
CH	1.09	1.05	0.91	0.96	1.16	1.00	1.00	1.00	0.92	1.08

3.3.2.2 Technical change

Having analyzed the efficiency changes for all these countries, we now take into account their changes in the frontier technology so as to fully evaluate the total factor productivity change of each country. The results are shown in Table 3.4.

Table 3.4 Technical changes of the 28 European countries from 2000 to 2009

Country	TECHCH									
	01/00	02/01	03/02	04/03	05/04	06/05	07/06	08/07	09/08	09/00
BE	1.02	1.03	1.05	1.12	1.09	1.04	1.02	1.10	1.13	1.78
BG	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
CZ	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
DK	1.00	1.01	1.02	1.19	1.07	1.02	1.04	1.06	1.10	1.62
DE	1.02	1.03	1.05	1.13	1.09	1.09	0.99	1.09	1.12	1.79
EE	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
IE	1.00	1.01	1.02	1.13	1.11	0.98	1.05	1.10	1.14	1.64
EL	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
ES	1.01	1.03	1.03	1.14	1.09	1.07	1.02	1.07	1.12	1.76
FR	1.02	1.03	1.04	1.11	1.10	1.01	1.02	1.12	1.14	1.75
IT	1.02	1.03	1.05	1.13	1.09	1.11	0.98	1.09	1.12	1.80
CY	1.00	1.01	1.02	1.19	1.09	1.07	1.01	1.08	1.12	1.73
LV	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
LT	1.00	1.01	1.02	1.20	1.09	0.98	1.05	1.11	1.13	1.72
LU	1.02	1.03	1.05	1.13	1.09	1.12	0.98	1.08	1.12	1.79
HU	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
NL	1.00	1.01	1.02	1.20	1.08	1.04	1.04	1.06	1.09	1.65
AT	1.02	1.03	1.05	1.13	1.09	1.10	0.99	1.08	1.12	1.79
PL	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
PT	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
RO	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
SI	1.02	1.03	1.04	1.11	1.10	1.01	1.03	1.11	1.14	1.73
SK	1.00	1.01	1.02	1.20	1.08	1.03	1.03	1.05	1.09	1.61
FI	1.00	1.01	1.00	1.10	1.14	0.97	1.05	1.10	1.14	1.61
SE	1.02	1.03	1.05	1.11	1.10	1.00	1.03	1.11	1.13	1.74
UK	1.01	1.02	1.00	1.11	1.11	1.00	1.03	1.12	1.14	1.66
NO	1.06	0.99	1.01	1.10	1.15	0.94	1.07	1.06	1.14	1.62
CH	1.02	1.03	1.05	1.13	1.09	1.11	0.97	1.09	1.12	1.79

We can see from Table 3.4 that although fluctuations occurred in every country within these ten years, the overall technical changes of these 28 European countries were all greater than one and at least 60% improvement with respect to their technology performance has been achieved during the past decade (see the last column of Table 3.4). Among others, Italy, Switzerland, Austria, Luxembourg, Germany, and Belgium were the technological innovators, which recorded an improvement of around 80% compared to 2000.

3.3.2.3 Total factor productivity change

Considering both efficiency change and technical change together, the overall road safety progress in each of these 28 European countries during the last decade can now be deduced, which is illustrated in Figure 3.4. Except for Romania, which had an overall *MI* value less than one indicating deterioration in its road safety productivity, all other countries have improved their road safety performance during this period, which is consistent with the fatality change as presented in Figure 3.1. Among others, Spain was the best performer, and seven countries, including Spain, Latvia, Cyprus, Portugal, Slovenia, Ireland, and Estonia, have already doubled their road safety performance compared with that in 2000. Luxembourg, France, Switzerland, and Germany also nearly made it with an improvement of above 90%, and they will possibly catch up with those better-performing ones by 2010 at the current rate of advance. The remaining countries have progressed however to a lesser extent, especially in Bulgaria, Poland, Hungary, and Greece, less than 50% improvement has been achieved during these ten years. Therefore, great efforts are still needed in such countries.

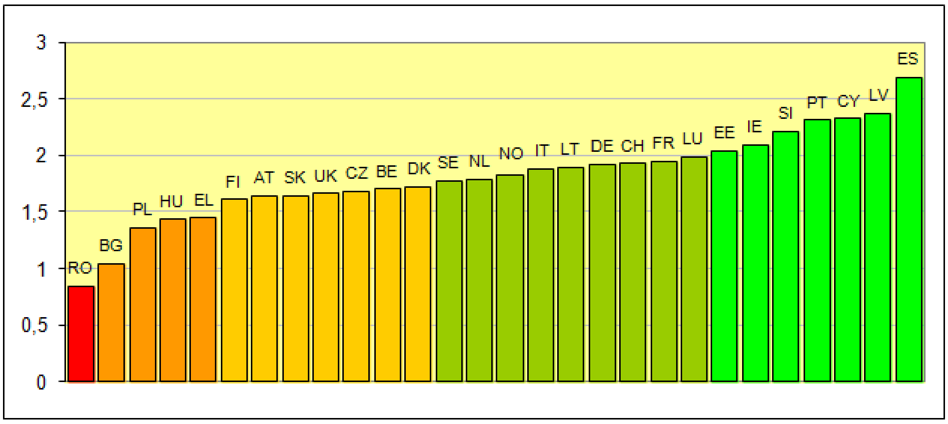


Figure 3.4 Overall road safety progress in the 28 European countries from 2000 to 2009

Further comparing the result with the one in Figure 3.1, which is only based on the fatality change between 2000 and 2009, we can see that although Latvia achieved the highest reduction in road fatalities (i.e., 60%), its overall road safety performance change was inferior to Spain based on the DEA-RS-MI, which could be mainly attributable to the reduction in its population size in the last decade (see Table 3.1). In other words, the great progress in the number of road fatalities in Latvia was partially offset by the reduction of its exposure during the same time period. On the contrary, due to the prominent reduction in road fatalities and rapid growth in the degree of participation in traffic as well, Spain actually achieved the highest road safety progress among all these 28 European countries in this period (see also the evolution in *EFFCH*, *TECHCH*, and *MI* of Spain in Appendix II). The same goes for Cyprus, Slovenia, Ireland, and so on. All these verify the fact that simply considering the reduction of the final outcome may not correctly reflect the real improvement in road safety performance because the transport circumstances of a country which can impact on the final outcome also changes every year. The approach used in this study makes the comparisons between countries more justly.

3.4 Conclusion

By using the DEA-based road safety model and the Malmquist productivity index, this chapter presented a new way for assessing the road safety performance change of countries over time. In doing so, we not only focused on the evolution of road safety final outcomes within a given period, but also took the changes in exposure in the same period into account. More specifically, using the information on the three measures of exposure, i.e., the number of inhabitants, passenger-kilometres travelled and passenger cars on the one hand, and the number of fatalities in road transport on the other hand, the Malmquist productivity index based on the DEA-based road safety model has proven valuable as a benchmarking tool for measuring the extent to which the 28 European countries have improved their 'productivity' on road safety over the period 2000-2009, and it has provided more objective results than the ones based on the traditional indicator only measuring the percentage change in road fatalities. The analysis found that there was a significant road safety progress in Europe during the last decade. However, the development in different countries was unbalanced. Some of them were still getting stuck in the rut or even deteriorating in terms of their road safety performance. Moreover, the decomposition of the DEA-RS-MI into technical change and efficiency change further revealed that the bulk of the improvement was attained through an overall improvement in the technological environment, rather than through the relatively inefficient countries catching up with those efficient ones. In the future, explorations on the reasons behind the progress or decline in each country and the prediction for its future development are worthwhile. In doing so, however, detailed information (to be determined, e.g., by means of interviewing national experts) is required.

Chapter 4 Serious Injuries: An Additional Indicator for Road Safety Evaluation

This chapter illuminates the impact of including serious injuries in addition to the fatalities in the road safety product benchmarking, thereby corresponding to the fourth research question of this dissertation. In doing so, different types of weight restrictions are formulated in the DEA-based road safety model to indicate their relationship.

4.1 Introduction

In the previous chapters of this thematic part, the number of road fatalities was used as the only road safety final outcome to benchmark the road safety performance and development of a set of European countries. This is not because road fatalities are the only interest but mainly because there is no reliable reporting or even the same definition on the number of crashes and the range of injury severities in different countries. As a result of this lack of comparability, most of the road safety studies for cross-country analysis (e.g., Al-Haji, 2007; Traynor, 2008, 2009; Gaygısız, 2010) focus entirely on fatalities, which however, represent only the 'tip of the iceberg' of the road crash problem and could lead up to an overestimation of this aspect. Consequently, it is highly desirable to extend the inter-national comparisons of road safety by taking a larger picture of the impact of road crashes into account.

In recent years, great efforts are being made in Europe to accomplish the harmonization on common definitions of injury severity and also its reporting procedures for the purpose of more complete inter-national benchmarking of road safety [European Transport Safety Council, 2008; Organization for Economic Co-operation and Development/International Transport Forum, 2011b]. In this study, an initial attempt of including the number of serious injuries as an additional indicator for road safety product benchmarking is carried out. We give priority to the level of serious injuries because of its greater impacts on society.

More importantly, a similar definition of serious injuries has already been applied in some European countries and they are more likely to be reported to the police than slight injuries and property-damage-only crashes. More information on the serious injuries in Europe is presented in Section 4.2. Methodologically, to integrate serious injuries with fatalities in a road safety benchmarking study, apart from using the DEA-based road safety model, additional weight restrictions are needed to indicate their relationship. Different types of possible weight restrictions are therefore elaborated in Section 4.3. Specific models for this study are demonstrated in Section 4.4, and the results are compared subsequently to show the impact of including serious injuries in the road safety product benchmarking for the countries under consideration. Finally, the concluding remarks are given in Section 4.5.

4.2 Serious Injuries in Europe

Today, due in large part to reinforcement of road user training and education, to advances in vehicle and infrastructure design and technology, as well as to improvements in medical care (e.g., prompt emergency response, early diagnosis, and treatment capabilities), many road fatalities are prevented and the downward trend is likely to continue in Europe. However, many survivors remain seriously injured, and their physiological and psychological consequences may last for days, months, years or even the rest of their life. In 2010, apart from the 31,000 people killed in road traffic crashes in the EU, police reports mention more than 340,000 people that were seriously injured[10] [European Transport Safety Council, 2011]. For each road fatality on Europe's roads, at least 44 injuries are recorded, of which 8 are categorized as ‘serious’ [European Transport Safety Council, 2010]. In other words, road fatalities form only a small minority of the totals, whereas non-fatal injuries, especially those seriously injured, are of importance in terms of both social and economic costs, and now represent an increasing concern of public authorities [European Transport Safety Council, 2007].

[10] The number of seriously injured persons recorded in hospital data is much larger, and the difference is, e.g., estimated as a factor of about 2 in countries such as the SUN countries [Broughton et al., 2008; European Transport Safety Council, 2010].

At the international level in general and in Europe in particular, however, it is not yet possible to make comparisons between all the Member States with respect to their number of serious injuries because both the level of injury reporting and the national definition of a serious injury vary greatly among countries [Broughton et al., 2008]. At present, only 16 EU countries, i.e., Belgium, Cyprus, Czech Republic, Denmark, France, Germany, Greece, Ireland, Latvia, Luxembourg, the Netherlands, Portugal, Slovakia, Spain, Sweden and United Kingdom, as well as Switzerland, use the same definition about serious injuries, which is 'spending at least one night in hospital as an in-patient or a close variant of this' [European Transport Safety Council, 2010]. Despite the fact that the level of reporting could still be different under the same definition due to differences in legislation, insurance policy, police resources and quality of data collection and processing in different countries, it provides at least a common basis to integrate this measure together with the fatalities for the road safety benchmarking of countries. Figure 4.1 illustrates the evolution in the number of serious injuries recorded in these countries[11] from 2001 to 2008, together with the fatality change during the same time period.

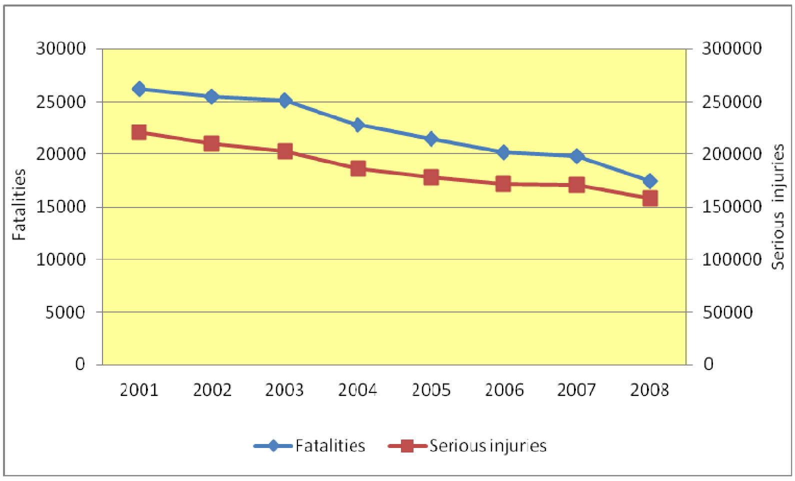


Figure 4.1 Evolution in road fatalities and serious injuries in 15 European countries over the period 2001 to 2008

[11] Latvia and France are not included because they use the same definition on serious injuries only since 2004 and 2005, respectively.

Although the number of seriously injured road users registered in this group of European countries was reduced by 28% during this period, the absolute value was still much higher than the number of recorded fatalities and its progress was also slower than that of fatalities (33%). Moreover, the situation differs considerably from country to country (see Appendix III). For instance, Portugal and Ireland have reduced their number of serious injuries by more than half during this period, which was faster than the reduction of their fatalities; Countries like Spain and United Kingdom have made progress on their serious injuries at a similar pace as their fatalities; Whereas in Luxembourg, more serious injuries were recorded in 2008 compared to 2001, although its fatality number has been reduced by half over this period. Consequently, to evaluate a country's road safety performance and to make comparisons with others, it is not correct to neglect this less-publicized part of the real picture by referring only to road fatalities.

4.3 Weight Restrictions in DEA

To integrate serious injuries with fatalities for road safety benchmarking purposes, the DEA-based road safety model (see Section 2.4.1) re-presented in (4-1) can be easily applied due to its powerful capability of handling multiple inputs and multiple outputs simultaneously.

$$
\begin{aligned}
\max E_0 &= \sum_{i=1}^{m} v_i x_{i0} \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{r0} = 1, \\
& \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \le 0, \quad j = 1, \cdots, n \\
& u_r, v_i \ge 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m
\end{aligned}
\tag{4-1}
$$

However, when computing the efficiency in DEA, apart from the non-negativity of the weights (i.e., $u_r, v_i \ge 0$), the model allows the weights associated with each input and output to be freely estimated in order to maximize the relative efficiency score of the DMU under consideration. The flexible allocation of input and output weights is often presented as advantageous in its applications since an a priori specification of the weights is not required and each DMU is evaluated in its best possible light. Thereby, if a DMU turns out to be inefficient, its inefficiency cannot be traced back to an inappropriate evaluation process [Vercellis, 2009]. Nevertheless, there are also disadvantages to this complete flexibility. Specifically, an unreasonable weight scheme could happen in which some DMUs would heavily weigh a few favorable inputs and outputs and completely ignore others in order to achieve a high relative efficiency score. One then faces the risk of basing global performance of a DMU on only a small subset of its factors. Moreover, since the weights derived from the model may vary a lot from one DMU to another, they can be in conflict with a priori knowledge or accepted views on the relative weights or rates of the factors.

One way to limit the range of values that the weights can take is to use weight restrictions. They could be imposed based on a priori knowledge about the weights or on the value judgments from experts, and can be incorporated in the multiplier DEA model directly. A large diversity of weight restriction techniques have been proposed in the DEA literature and their classification is also well documented (e.g., Dyson & Thannassoulis, 1988; Wong & Beasley, 1990; Allen et al., 1997; Thanassoulis et al., 2004; Cherchye et al., 2007a). In the following sections, some commonly used types of weight restrictions are outlined, and their implications in the DEA modeling framework are interpreted.

4.3.1 Absolute weight restrictions

This type of weight restrictions assigns upper and/or lower bounds on the absolute values of the input or output weights, which is illustrated by (4-2).

$$\begin{aligned} &LI_i \le v_i \le UI_i \quad i = 1,\cdots,m \\ &LO_r \le u_r \le UO_r \quad r = 1,\cdots,s \end{aligned} \tag{4-2}$$

The bounds used in the restrictions are mainly introduced to prevent the corresponding inputs or outputs from being overemphasized or ignored in the analysis. However, there is a strong interdependence between the bounds on different weights. For instance, setting an upper bound on one output weight means imposing a lower bound on the total weighted output of the remaining variables and this in turn has implications for the values that the remaining output weights can take. In general, the efficiency score of a DMU is worsened due to the additions of these restrictions and they may also render the model infeasible. Hence, careful attention is needed in determining these bounds. In this respect, recourse to auxiliary information such as shadow prices[12], unit costs, etc, is often used.

4.3.2 Relative weight restrictions

The second type of weight restrictions is depicted in (4-3), which is also named the *assurance region* method [Thompson et al., 1990]. Different from assigning restrictions on the absolute values of input or output weights, this type of restriction is introduced to incorporate into the analysis the relative values of inputs and/or outputs weights that vary between an upper and/or lower bound.

$$\begin{aligned}
&LI_{i,i+k} \le \frac{v_i}{v_{i+k}} \le UI_{i,i+k} \quad i=1,\cdots,m-1, \quad k=1,\cdots,m-i \\
&LO_{r,r+k} \le \frac{u_r}{u_{r+k}} \le UO_{r,r+k} \quad r=1,\cdots,s-1, \quad k=1,\cdots,s-r \\
&LIO_{i,r} \le \frac{v_i}{u_r} \le UIO_{i,r} \quad i=1,\cdots,m, \quad r=1,\cdots,s
\end{aligned} \tag{4-3}$$

Relative weight restrictions are particularly suitable when translating a priori knowledge or expert opinions on the pairwise relation of the factors. They would thus capture requirements or statements such as 'the price/size of output *X* can at most be double the one of output *Y*'.

It should be noted that the bound values for relative weight restrictions are dependent on the scaling of the inputs and outputs, that is, they are sensitive to the measurement unit of the related factors. Moreover, a DMU previously characterized as efficient may be found to be inefficient after such restrictions are imposed, and they may also render the model infeasible.

4.3.3 Ordinal weight restrictions

Combining both absolute and relative weight restrictions, we obtain the third type, called ordinal weight restrictions, in which the weights of more than two

[12] The shadow price reflects the marginal rate of substitution between inputs and outputs, which measures the extra value that would come from increasing the most relevant production resource by one unit. In turn, this indicates the highest price the producer can pay for that added resource without becoming worse off overall [Kanbur, 1987].

factors are compared simultaneously in an ordinal manner. One possible form of this restriction type is shown as follows.

$$\begin{aligned} & LI_i \le v_i \le v_{i+1} \le \cdots \le v_{i+k} \le UI_{i+k} \quad i = 1,\cdots,m-1, \ \ k = 1,\cdots,m-i \\ & LO_r \le u_r \le u_{r+1} \le \cdots \le u_{r+k} \le UO_{r+k} \quad r = 1,\cdots,s-1, \ \ k = 1,\cdots,s-r \end{aligned} \qquad (4\text{-}4)$$

The usage of this type of weight restrictions is context dependent. It inherits all the properties of the above two restrictions but their cautions also need to be paid attention to.

4.3.4 Virtual weight restrictions

Rather than directly restricting the actual input and output weights introduced above, another widely used type of weight restrictions is to limit the value of the virtual inputs or outputs, i.e., the product of the input or output and its corresponding weight [Wong & Beasley, 1990]. Let us consider a virtual weight restriction on output *r* shown in (4-5):

$$LO_r \le \frac{u_r y_{rj}}{\sum_{r=1}^{s} u_r y_{rj}} \le UO_r \quad r = 1,\cdots,s \qquad (4\text{-}5)$$

the proportion of the total virtual output of DMU *j* devoted to output *r*, i.e., the 'importance share' attached to output *r* by DMU *j*, is restricted to range between [LO_r, UO_r], which have a value between 0 and 1, respectively. These restrictions are attractive in view of the fact that expert opinion is often collected by a budget allocation approach [Organization for Economic Co-operation and Development, 2008], in which experts are asked to distribute let's say 100 points over the different dimensions to indicate importance. A similar restriction can also be set on the virtual inputs.

$$LI_i \le \frac{v_i x_{ij}}{\sum_{i=1}^{m} v_i x_{ij}} \le UI_i \quad i = 1,\cdots,m \qquad (4\text{-}6)$$

(4-5) and (4-6) are actually the absolute virtual weight restrictions, and the relative and ordinal virtual weight restrictions can also be formulated accordingly. It is important to note that imposing such a weight restriction may introduce a comparison with non-existent targets. More information on these types of weight restrictions can be found in Sarrico & Dyson (2004) and Cherchye et al. (2007a).

4.3.5 Summary

In this section, several approaches to the use of weight restrictions in DEA have been presented. They are all application driven, so no overall approach can be identified as suitable for all different circumstances, and there are still some other techniques available for restricting the weights in DEA, such as the cone-ratio method [Charnes et al., 1990]. No matter which approach is applied in a specific context, the main purpose is to restrict the flexible selection of input and/or output weights in the basic DEA framework, and to guarantee the establishment of a proper weighting scheme. However, the key difficulty in using any of these weight restrictions outlined above is the estimation of appropriate values for the constants in the restrictions, compatible with a priori knowledge or the reflection of the value judgments from experts in the efficiency assessment. In this study, to exhibit a larger picture of the impact of road crashes by considering both the number of road fatalities and serious injuries, it is important to indicate their relationship so as to obtain reasonable benchmarking results. To this end, relative weight restrictions and virtual weight restrictions are applied respectively, and the results are discussed in the following section.

4.4 Application and Results

To benchmark the road safety performance of the aforementioned 17 European countries which have the same definition on both road fatalities and serious injuries, the DEA-based road safety model (4-1) is utilized. The three common measures of exposure to risk used in the previous Chapters, i.e., the number of inhabitants, passenger-kilometres travelled, and passenger cars, are still the model's inputs and the number of road fatalities and the number of serious injuries are the two outputs (see Figure 4.2). Data from 2006 to 2008 are collected from the European Commission (2011a) and the European Transport Safety Council (2010), and the average values (see Table 4.1) are used in the analysis so as to avoid coincidental fluctuation in the data and to improve the reliability of the results.

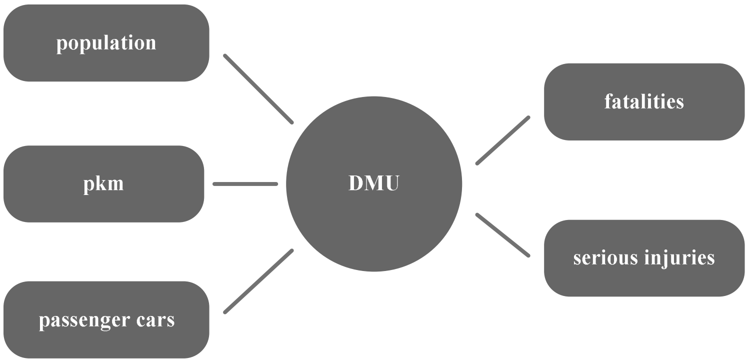


Figure 4.2 The input and output of the model

Table 4.1 Average input and output values of the 2006-2008 period for 17 European countries

Country	Input			Output	
	Population (million)	Passenger-kilometres (10 billion)	Passenger cars (million)	Fatalities	Serious injuries
BE	10.63	11.09	5.02	1028.00	7043.00
CY	0.78	0.54	0.39	85.67	702.67
CZ	10.34	7.12	4.19	1120.00	3823.00
DK	5.46	5.24	2.04	372.67	2960.00
FR	61.95	72.39	31.08	4534.67	38080.67
DE	82.25	86.12	41.06	4839.00	73529.67
EL	11.19	9.50	4.67	1608.00	1898.67
IE	4.35	4.72	1.84	327.33	802.33
LV	2.28	1.57	0.85	380.67	686.33
LU	0.48	0.66	0.32	38.00	289.33
NL	16.39	14.79	7.31	705.33	9348.00
PT	10.61	8.65	4.32	942.67	3068.33
SK	5.40	2.62	1.40	588.00	1958.00
ES	44.85	34.11	21.29	3675.67	19055.00
SE	9.15	9.82	4.23	437.67	3813.33
UK	60.99	68.20	28.67	3000.67	27859.33
CH	7.56	8.27	3.93	370.33	5027.00

Source: European Commission (2011a); *European Transport Safety Council (2010)*

Before applying the DEA-RS model, weight restrictions on the road fatalities and the serious injuries have to be specified in order to reflect their relationship and to guarantee reasonable results[13]. In this respect, information on their (relative) shadow price and a priori knowledge about the relative importance of these two aspects are possible recourses.

4.4.1 Relative weight restrictions based on shadow price

There has been much research relating to valuation (or shadow price) of fatal and non-fatal effects of accidents (e.g., European Conference of Ministers of Transport, 1998). Amongst others, O'Reilly et al. (1994) presented the results of a major UK research project on the valuation of serious injuries on the road. The main advantage of using these results is that the valuation of serious injuries was expressed as a ratio of that of fatalities (see Table 4.2), which is particularly suitable for imposing relative weight restrictions in this study.

Table 4.2 Ratios of the valuation of serious injuries and that of fatalities

Valuation method	Ratio
Standard gamble	8.6-12.2% (best=9.5%)
Contingent valuation	29.1-54.1% (best=37%)
Expert ranking	18-20%

Source: O'Reilly et al. (1994)

Three research techniques were utilized, in which the standard gamble method generated the lowest best ratio (9.5%) while the contingent valuation yielded the highest (37%). A third strand of research, using health experts' rankings of severity of different injury states and recovery times, alongside relative measurements of 'time lost', gave values in between the two other techniques. In this study, the above information (in particular the best ratios) is used to indicate the relationship between fatalities (y_1) and serious injuries (y_2). A

[13] Although the three inputs describe the exposure to risk in road transport from different perspectives, they are highly correlated. It is therefore not so necessary to impose a weight restriction for them.

relative weight restriction is thereby imposed and further incorporated in the DEA-RS model as follows:

$$
\begin{aligned}
\max E_0 &= \sum_{i=1}^{m} v_i x_{i0} \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{r0} = 1, \\
& \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \le 0, \quad j = 1, \cdots, n \\
& 0.095 \le \frac{u_2}{u_1} \le 0.37, \\
& u_r, v_i \ge 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m
\end{aligned} \tag{4-7}
$$

4.4.2 Virtual weight restrictions based on a priori knowledge

The shadow price offers valuable information to narrow the range of values that the weights of corresponding factors can take. However, the data are not always available and the values are also time-varying, which to a certain extent limits the use of this information for setting appropriate weight restrictions. In this study, an alternative way to reflect the relationship between the road fatalities and the serious injuries is to use a priori knowledge on these two aspects, i.e., fatalities play in most cases a more important role in determining the road safety performance of a country than serious injuries. A representative virtual weight restriction can then be imposed to DMU_0 as follows:

$$
\frac{u_2 y_{20}}{\sum_{r=1}^{2} u_r y_{r0}} < \frac{u_1 y_{10}}{\sum_{r=1}^{2} u_r y_{r0}} < 2 \frac{u_2 y_{20}}{\sum_{r=1}^{2} u_r y_{r0}} \tag{4-8}
$$

It indicates that fatalities account for a higher percentage share in road safety evaluation than serious injuries, but the fatality share is also restricted to be less than twice the share of the serious injuries to avoid irrational allocation. This restriction can then be incorporated into the DEA-RS model as below:

$$\begin{aligned}
\max E_0 &= \sum_{i=1}^{m} v_i x_{i0} \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{r0} = 1, \\
& \sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \le 0, \quad j = 1, \cdots, n \\
& u_2 y_{20} < u_1 y_{10} < 2u_2 y_{20}, \\
& u_r, v_i \ge 0, \quad r = 1, \cdots, s, \quad i = 1, \cdots, m
\end{aligned} \tag{4-9}$$

4.4.3 Result comparisons

Applying models (4-7) and (4-9), respectively, we obtain the overall road safety efficiency scores of the 17 European countries by taking into account both their road fatalities (F) and serious injuries (SI) as output in the benchmarking study. The scores are presented in Table 4.3, along with the ones derived from only considering the number of road fatalities using the DEA-RS model (4-1).

Table 4.3 Road safety efficiency scores of the 17 European countries based on different models

Country	F & SI Model (4-7)	Country	F & SI Model (4-9)	Country	Only F Model (4-1)
SE	1	SE	1	SE	1
IE	1	IE	1	UK	1
UK	0.983	UK	0.991	NL	1
LU	0.936	NL	0.984	CH	1
NL	0.899	CH	0.967	DE	0.808
CH	0.877	LU	0.930	LU	0.789
PT	0.789	EL	0.806	FR	0.705
ES	0.784	ES	0.784	DK	0.654
DK	0.734	PT	0.777	IE	0.640
FR	0.725	DE	0.754	ES	0.553
EL	0.695	DK	0.734	BE	0.487
DE	0.657	FR	0.724	PT	0.484
SK	0.640	CZ	0.630	CY	0.436
CZ	0.639	SK	0.629	CZ	0.397
BE	0.577	BE	0.574	SK	0.394
LV	0.515	LV	0.531	EL	0.299
CY	0.494	CY	0.492	LV	0.257

Comparing the results in the second and the fourth column of Table 4.3, which are obtained by considering both the road fatalities and the serious injuries as output for a more complete inter-national benchmarking of road safety, we find that no matter which weight restriction is applied, the country rankings are more or less consistent, especially for those best- and worst-performing countries. The high correlation coefficient between these two sets of efficiency scores (0.968) also implies the substantial equivalence of using either of these two restrictions. However, it should be noted that no feasible solution is found when these two weight restrictions are used simultaneously in this case, which means that some conflicts still exist between the two restrictions. One possibility is that the shadow price used in this study as estimation for the constants in the relative weight restriction may not be so precise. On the other hand, it is also possible that the knowledge on road fatalities and serious injuries used for setting the virtual weight restriction is not satisfiable for all countries. In other words, the price paid for the serious injuries in some countries may be higher than that for the fatalities, or the price paid for the fatalities may be two times higher than that for the serious injuries. In addition, the inaccuracy of the input and output data, especially the number of serious injuries, is also a possible reason for this infeasibility. All these in turn help in interpreting the differences of these countries in terms of their efficiency score and ranking based on the different models.

Moreover, to illuminate the impact of including serious injuries as an additional indicator of road safety final outcome, we compare the results with the ones based on only road fatalities, which are presented in the last column of Table 4.3. It can be seen that Sweden is the only best-performing country no matter which model is utilized. It indicates the outstanding performance of Sweden regarding both road fatalities and serious injuries. However, countries such as United Kingdom, the Netherlands and Switzerland, which obtain the efficiency score of one when only considering the number of fatalities in the road safety benchmarking, are no longer efficient when the serious injuries are integrated. It implies that the situation of serious injuries in these countries is still serious. Including this aspect deteriorates to a certain extent the overall road safety performance of these countries. The same situation also applies to countries like Germany, France, Belgium, and Cyprus. On the contrary, Ireland, which ranks in the middle among all the countries when only the number of road fatalities is considered, becomes one of the best-performing countries when integrating its number of serious injuries in the evaluation. It means that the situation of serious injuries in Ireland is much better than that of other countries, which enables this country to become one of the valuable benchmarks for other countries to learn from, especially concerning its best practice on reducing the number of serious injuries. Apart from Ireland, the ranking of countries such as Greece, Portugal, and Spain is also improved when the serious injuries are considered[14]. In general, no matter whether the country's road safety ranking is improved or deteriorated, most of the underperforming countries achieve a higher efficiency score when the number of serious injuries is included, even the worst-performing country, i.e., Cyprus. In other words, the distance between those best-performing and underperforming countries becomes smaller when a larger picture of the impact of road crashes is taken into account. In a sense, it represents a more realistic relative situation of road safety between these countries[15].

4.5 Conclusion

In this chapter, we investigated the possibility of including the number of serious injuries as an additional indicator of road safety final outcome to perform road safety product benchmarking and further illuminated its impact on the countries' ranking. In doing so, the DEA-based road safety model was utilized, and additional weight restrictions were introduced to indicate the relationship between road fatalities and serious injuries. In this study, a relative weight restriction based on the information of their valuation ratios from literature and a virtual weight restriction using a priori knowledge on these two aspects were incorporated in the model respectively to compute the efficiency score of 17

[14] It has to be noted that the improvement in ranking of these countries might be partly influenced by their different levels of underreporting with respect to the number of serious injuries.

[15] Better-performing countries are more likely to prevent fatalities from serious injuries. Therefore, a larger distance between best-performing and underperforming countries is expected when only the number of fatalities is considered.

European countries which have the same definition on both road fatalities and serious injuries. The results indicated the substantial equivalence of using these two different restrictions. Moreover, comparing the results with the ones from only considering the number of road fatalities, the impact of including the number of serious injuries as an additional indicator of road safety final outcome was discussed. In general, representing a larger picture of the impact of road crashes affected the ranking of the countries to some extent. Moreover, no matter whether the country's ranking was improved or deteriorated, a higher efficiency score was achieved by most of the underperforming countries. Given the importance of considering the serious injuries in addition to the fatalities for inter-national benchmarking of road safety, the proposed model (i.e., the DEA-RS model with weight restrictions) turned out to be effective in deriving reasonable results. We are thereby also inspired to apply this kind of model to a more complete road safety product benchmarking practice in the future when the data on for example the number of crashes, the degree of property damage, and the number of slight injuries are ready to use.

Finally, although the methodology is well established to include more indicators for road safety product benchmarking, we should still keep in mind that the number of serious injuries is not yet a mature indicator due to large differences in reporting practices in different countries (see also [Organization for Economic Co-operation and Development/International Transport Forum, 2011]). International cooperation in terms of injury data collection and harmonization is therefore sorely required, and further efforts to link police reports to other data sources (e.g., hospital records) are also essential to improve data quality and consistency. They are the fundamental condition of making comparisons between countries, and also the key to designing more effective safety policies.

Part II Towards a Composite Road Safety Performance Index

Introduction to Part II

The first thematic part of this thesis concentrated on road safety product benchmarking, in which different road safety final outcomes were investigated and the corresponding risk indicators based on different measures of exposure as well as their evolution were compared between countries. Useful benchmarks were therefore identified and practical targets in terms of road fatalities were assigned for those underperforming countries. However, setting targets does not guarantee their achievement. Road safety policy makers and analysts aiming at a higher level of safety or a lower number of final outcomes need to take into account as many factors influencing safety as possible or, at least, those factors they are able to affect or control. To this end, a second type of benchmarking study, i.e., road safety programme benchmarking, which is used to compare human-vehicle-infrastructure performance between countries, has received considerable policy attention nowadays, and is the main focus of this thematic part as well.

In doing so, safety performance indicators (SPIs) – which are causally related to the number of crashes or to the injury consequences of a crash – such as levels of mean traffic speeds, seat belt wearing, drink driving, vehicle and road safety ratings, etc., have to be developed firstly. Moreover, to measure the multi-dimensional concept of road safety performance which cannot be captured by a single indicator, the exploration of a composite road safety performance index is vital for rational benchmarking of road safety. However, the task of constructing such a composite index raises a number of research issues, some of which have not yet been properly addressed in current road safety studies.

First, since more and more SPIs are developed and increasingly used to comprehensively quantify the entire situation of possible risk factors, they are much likely to be grouped into different categories and further be linked to one another constituting a layered hierarchy. It provides a detailed insight into the structure of indicators and is worthwhile to be reflected in the index construction. However, it is prone to be ignored in the current index research partly due to the limitation of traditional weighting and aggregation techniques in reflecting this kind of hierarchical structures.

Coupled with the proliferation of SPIs, some practical issues related to data also inevitably emerge in the development of a road safety performance index, two of which are qualitative indicators and missing values. Specifically, obtainment of measurable and quantitative indicators is commonly the prerequisite of any index research. This, however, becomes more and more difficult to be guaranteed since the natural uncertainty of reality often leads up to imprecision and vagueness inherent in the information that can only be represented by means of qualitative indicators. Simply treating them as quantitative ones could thereby result in wrong conclusions.

Moreover, an extension of the data set used for road safety index research raises the issue of missing values, which to a great extent restricts researchers from performing classical analyses as complete data matrices are usually required. Consequently, how to effectively tackle these data problems directly affects the result of the road safety index research and the success of the benchmarking as well.

Taking into account the aforementioned research challenges which have not yet been systematically investigated in the current road safety index studies, the following four research questions are highlighted as the main focus of this part of the thesis:

RQ5: Which are current available national safety performance indicators and how can they best be structured?

RQ6: How to reflect a layered hierarchy of indicators in constructing a road safety performance index and what is the added value?

RQ7: How to obtain a reliable index score for each country when missing data exist?

RQ8: What is the possible way to incorporate qualitative indicators?

Chapter 5 Development of Safety Performance Indicators and Data Processing

This chapter identifies the current available national safety performance indicators that are valuable to be used for inter-national programme benchmarking of road safety. Their hierarchical structure is established, corresponding data are collected, and necessary data processing is performed. This chapter thereby corresponds to the fifth research question of this dissertation.

5.1 Introduction

To obtain a composite road safety performance index for the sake of meaningful inter-national road safety programme benchmarking, a comprehensive set of indicators that corresponds to as many underlying risk factors influencing safety as possible has to be developed in the first place. In this respect, safety performance indicators (see the road safety target hierarchy in Figure 1.6), which are viewed as intermediate outcomes (such as the proportion of car occupants using seat belts) linking safety countermeasures (such as the installation of seat belt reminders in passenger cars) with final outcomes (such as casualties in road crashes), are widely investigated in current road safety studies.

In contrast to crash data, which are frequently treated as the 'worst case scenario' in the unsafe operational conditions of the traffic system and are insufficient in explaining more detailed aspects of crash causation and injury prevention, SPIs are defined as any measurement that is causally related to the number of crashes or to the injury consequences of a crash, and are used in addition to the figures of crashes or injuries in order to indicate safety performance or understand the process that leads to crashes [European Transport Safety Council, 2001]. The purpose of SPIs is threefold: to reflect the current safety conditions of a road traffic system; to measure the influence of various safety interventions; and to compare between different road traffic systems such as countries or regions [Vis, 2005].

Due to the high information density, SPIs allow quicker and more local analyses and monitoring than crash data do. As believed, SPIs can give a more complete picture of the level of road safety and can point to the emergence of developing problems at an early stage, before these problems show up in the form of crashes [European Transport Safety Council, 2001]. Moreover, by linking the casualties from road crashes and the measures to reduce them, SPIs provide a means for monitoring the effectiveness of safety actions applied and for further guiding policy decisions regarding existing and new countermeasures.

Today, having recognized the complex character of the road safety phenomenon, a large number of factors involved in road safety development have been identified, more and more SPIs are thereby developed and increasingly used as a supportive instrument for inter-national comparisons of road safety performance, especially over the last decade (e.g., European Transport Safety Council, 2001; Vis, 2005; Al-Haji, 2007; Wegman et al., 2008; Hermans, 2009a; Gitelman et al., 2010).

In this chapter, the current available national SPIs are discussed based on the identification of various risk factors in road transport. Their hierarchical structure is then established, and corresponding data are collected (Section 5.2). Moreover, some necessary data processing procedures are provided with a view to the following index construction (Section 5.3). The chapter closes with conclusions in Section 5.4.

5.2 Indicator Development and Data Collection

Road safety problems have traditionally been viewed as the result of malfunctions in the road transport system, which consists of three main components: the road user, the vehicle and the road [World Health Organization, 2004]. Each crash is in most cases a direct consequence of failure in one or several of these three factors who influence each other (see Figure 5.1). As a result, the European Transport Safety Council (2001) recommended the development of SPIs related to *road user behavior*, *vehicle*, and *road*. In addition, trauma management, or *emergency medical services (EMS)* in particular, which is concerned with the medical treatment of injuries resulting from road crashes, was also highlighted in the report due to its significant influence on post-crash injury outcomes, and it often constitutes the factor of *infrastructure* together with the *road* (see Section 5.2.3). On the basis of this report, the European *SafetyNet* project [Hakkert et al., 2007a, b] provided a methodological basis for indicator development, and Hermans (2009a) further summarized from literature eight criteria for indicator selection, which are: relevant, measurable, understandable, data available, reliable, comparable, specific, and sensitive.

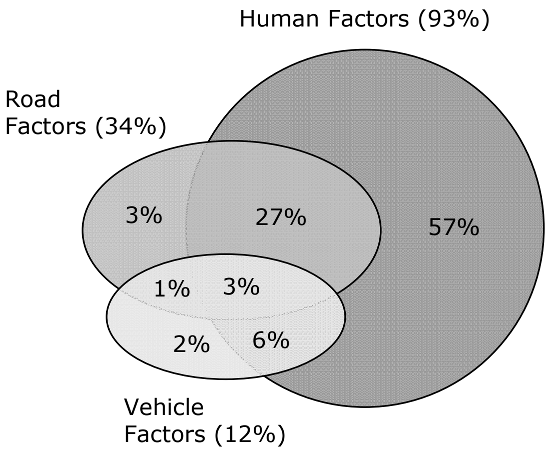


Figure 5.1 Venn diagram on crash factors

Source: Rumar (1985)

Specifically, SPIs that are developed for a certain safety component should reflect the factors contributing to road crashes/injuries and characterize the scope of the problem identified. The development of SPIs begins with a definition of the problem (i.e., the operational conditions of the road traffic system which are unsafe and result in crashes/injuries as the 'worst case') and continues with the conversion of this information into measurable variables. One example is ‘the proportion of car occupants using seat belts’. This SPI therefore represents a specific safety aspect (seat belt usage) as well as a comparable value (proportion of car occupants) of how this aspect has penetrated to the traffic system. Implicitly, the SPI should have a proven and well-documented relation to the number of casualties, and could be seen as an intermediate measurement of the safety level for that specific aspect. For the majority of SPIs, there are several countermeasures that could contribute to their improvement. Taking the above indicator related to seat belt usage as an example, the improvement could for instance follow as a result from seat belt legislation and enforcement, a demerit point system, or intelligent seat belt reminders in isolation or in combination.

In Europe, several initiatives and research studies have been implemented in order to assess the performance across the EU concerning particular road safety aspects, such as the *SARTRE* study (focuses on the road user behavior in the EU) [SARTRE 3 consortium, 2004], *EuroNCAP* study (focuses on safer vehicles in the EU) [Lie & Tingvall, 2000], and the *EuroRAP* study (focuses on safer roads in the EU) [Lynam et al., 2004]. Data are available in a number of European institutions and databases such as Eurostat, International Road Traffic and Accidents Database (IRTAD), European Transport Safety Council (ETSC), and European Road Federation (ERF). The inter-national best practice reviews indicate that despite some differences in levels of motorization, the road safety problems in most Member States have many similarities [European Transport Safety Council, 2001]. A number of common road safety risk factors are therefore designated as central to road safety activities in Europe and were selected for the development of SPIs as well [e.g., Hakkert et al., 2007a, b; Hermans, 2009a]. They are alcohol, speed, protective systems, vehicle, road, and emergency medical services. In the following sections, the importance of these risk factors in road safety is discussed, the best available SPIs for each risk factor are developed, and corresponding data are collected.

5.2.1 Road user behavior

Comprehensive studies on road safety (e.g., Treat et al., 1977; Rumar, 1982; Green & Senders, 2004) found that inappropriate road user behavior was the major contributory factor to road crashes. They indicated that human error was the sole cause in 57% of all crashes and was a contributing factor in over 90% (see also Figure 5.1). In Europe, the incidence of drinking and driving, speeding behavior, and the nonuse of various protective systems are recognized as the three most important risk aspects in terms of road user behavior, and are therefore used as the basis for the development of SPIs in this study.

5.2.1.1 Alcohol

Driving under the influence of alcohol is believed to increase the risk and severity of road crashes more than most other traffic law violations [Hakkert et al., 2007a]. Hakkert & Braimaister (2002) provided a review of many studies and reported that the risk in traffic would increase rapidly with blood alcohol concentration (BAC). The relative crash risk starts increasing significantly at a BAC level of 0.4g/l [World Health Organization, 2004]. A study from the United States [Zador, 1991] showed that for single-vehicle crashes, each 0.02% increase in BAC level nearly doubles the risk of getting involved in a fatal crash. In Europe, driving with excess alcohol is responsible for at least 20% of the serious and fatal injuries [European Transport Safety Council, 2001]. Therefore, a reduction in drink driving above the legal limit would make a large contribution to the improvement of road safety.

To compare the situation of drink driving between countries, the ideal SPI would be the prevalence and concentration of impairing substances among the general road user population in each country [Hakkert et al., 2007a]. However, several methodological problems exist associated with this SPI such that ensuring a national random sample of the road user population is difficult and costly. As a result, a less ideal but more practical SPI, i.e., *the percentage of drivers above the legal BAC limit in roadside checks*, is selected for substitution, which can be computed based on the total number of roadside alcohol breath tests and the number of positive alcohol (i.e., with BAC above the legal limit) among tests[16]. The indicator data (average value of 2006-2008) collected from the European Transport Safety Council (2010) are shown in the second column of Table 5.1.

As can be seen, only around half of these 28 European countries have data on this SPI (some of them only have one or two years data), while the other half either have not yet established such a data collection system, or even prohibit the random testing of drivers by the constitution such as in Germany.

[16] This indicator is less ideal also because not all the countries are currently using the same BAC limit.

Table 5.1 Alcohol related SPIs

	% of drivers above the legal BAC limit in roadside checks	Remark	% of fatalities attributed to alcohol	Remark
AT	7.40%		7.67%	
BE	N/A		5.43%	
BG	N/A		4.15%	
CY	6.29%		19.53%	
CZ	N/A		5.26%	
DK	N/A		24.78%	
EE	1.00%		44.44%	
FI	1.45%	mean of 07-08	26.01%	
FR	3.27%		28.86%	
DE	N/A		11.62%	
EL	3.15%		8.22%	
HU	3.08%		12.55%	
IE	3.63%	mean of 07-08	29.80%	2005
IT	N/A		3.58%	
LV	N/A		20.24%	
LT	1.55%		11.07%	
LU	N/A		14.29%	2004
NL	N/A		3.55%	
NO	N/A		22.32%	2005
PL	9.50%	2008	8.11%	
PT	6.29%		5.82%	
RO	N/A		8.40%	
SK	N/A		5.85%	
SI	7.04%		45.64%	
ES	2.16%		8.82%	
SE	0.86%		9.95%	
CH	N/A		15.42%	
UK	16.89%	mean of 06-07	15.55%	

Source: European Transport Safety Council (2010)

Under this circumstance, a second indicator, i.e., *the percentage of road fatalities attributed to alcohol*, which represents the consequence of drink driving from the view of the final outcome level, is used as a substitutive SPI because most countries supposedly test a large part of the drivers involved in fatal crashes for alcohol. The average value of 2006-2008 for this indicator is presented in the fourth column of Table 5.1. As can be seen, most of the countries have available data which range from 3.55% in the Netherlands to

45.64% in Slovenia. There are three countries (i.e., Ireland, Luxembourg, and Norway), which have no data for these three years. The last observed value is therefore used for substitution, such as 2005 data for Ireland. The same treatment goes to all the following indicators.

In addition to the above two SPIs, one more indicator related to policy output can also be used to supplement the alcohol performance of a country, which is *the effectiveness of the overall enforcement against drinking and driving*. Such a policy performance indicator, however, is qualitative in nature, and can only take the form of ordered classes rated on for instance a 0-10 scale rather than numerical values for the purpose of description, comparison and evaluation of this risk factor. How to deal with this kind of qualitative indicators will be elaborated in Chapter 8.

5.2.1.2 Speed

Apart from drink driving, speed is another main cause of road crashes and crash injuries, and hence, a major issue for road safety. Inappropriate or excessive speed has been recognized as one of the most important risk factors influencing both the number of road crashes and the severity of injuries [Elvik, 2005; Kweon et al., 2005]. Some speed-crash studies indicated that the probability that a crash will result in injury is proportional to the square of the speed; for serious injury proportional to the cube of the speed; and for fatal injury proportional to the fourth power of the speed [World Health Organization, 2004]. In around one third of the fatal crashes, speed is an essential contributory factor [Bowie & Walz, 1994; Transportation Research Board, 1998]. Therefore, reducing vehicle speeds appears to have a significant effect on road safety final outcomes. On average, a 1% reduction in the mean speed of traffic leads to a 2% reduction in crashes resulting in injuries, a 3% reduction in crashes resulting in severe injuries and a 4% reduction in fatal crashes. Moreover, Taylor et al. (2000) showed that the road crash risk increases with the proportion of drivers over the speed limit. The crash risk grows by 10% if the proportion of offenders doubles. In European countries, *the mean speed* and the level of compliance of vehicles (or *the proportion of vehicles exceeding the speed limit*) in free-flowing traffic are therefore the two most commonly used speed SPIs. Furthermore, since the risk linked to speed varies across road types, differentiation among motorways, rural roads and urban roads is considered when making comparisons between countries of their levels of speed and speed limit violations. The average indicator values of 2006-2008 are collected from the European Transport Safety Council (2010) and Vis & Eksler (2008) and are shown in Table 5.2.

Table 5.2 Speed related SPIs

	Mean speed*			% of speed limit violations		
	on motorways	on rural roads	on urban roads	on motorways	on rural roads	on urban roads
	06-08	06-08	06-08	06-08	06-08	06-08
AT	0.9103	0.8960	1.0307	21.33%	19.43%	53.87%
BE	1.0092	0.9426	1.0767	N/A	34.10%	61.53%
BG	0.9315	N/A	N/A	N/A	N/A	N/A
CY	1.0500	1.1000	0.9600	52.50%	55.00%	N/A
CZ	0.8346	0.7593	0.8800	75.00%	15.10%	24.30%
DK	0.9359	1.0558	1.0387	31.50%	69.77%	60.00%
EE	N/A	1.0544	N/A	N/A	24.90%	N/A
FI	0.8861	0.9600	N/A	39.90%	43.92%	N/A
FR	0.9154	0.8900	0.9867	32.33%	27.27%	42.97%
DE	N/A	N/A	N/A	N/A	N/A	N/A
EL	N/A	N/A	N/A	N/A	N/A	N/A
HU	0.8582	0.8830	1.0120	45.07%	30.07%	59.40%
IE	0.9028	0.9185	1.1420	16.33%	31.33%	61.33%
IT	N/A	N/A	N/A	N/A	N/A	N/A
LV	N/A	1.0174	N/A	N/A	50.90%	N/A
LT	0.8538	0.9815	1.1580	20.70%	39.35%	43.00%
LU	0.8846	N/A	N/A	5.00%	N/A	N/A
NL	0.9500	N/A	N/A	36.00%	N/A	N/A
NO	1.0000	0.9875	1.0453	51.50%	44.80%	N/A
PL	N/A	1.0044	1.2780	N/A	65.83%	82.63%
PT	1.0083	1.1333	0.9000	54.00%	74.00%	38.00%
RO	N/A	N/A	N/A	N/A	N/A	N/A
SK	N/A	N/A	N/A	N/A	N/A	N/A
SI	0.8846	0.7000	1.1600	34.00%	1.00%	84.00%
ES	0.9528	N/A	N/A	37.93%	N/A	N/A
SE	0.9664	0.9767	0.9500	67.80%	52.60%	52.80%
CH	0.9056	0.9250	0.8267	23.00%	23.33%	13.67%
UK	0.9941	0.7969	1.0063	51.67%	10.00%	49.33%

* To make data comparable between countries, the mean speed on each road type is normalized by the corresponding speed limit on that road type.

Source: European Transport Safety Council (2010) and Vis & Eksler (2008)

There are still a large number of missing values in the data set. Some countries even don't collect (or have) any of these six indicator values, such as Germany, Greece, Italy, Romania, and Slovakia. Based on the data available, we can see that countries are more likely to have a mean speed above the speed limit on their urban roads (with the mean speed value higher than one), which corresponds to the fact that most of the countries have a relatively higher percentage of speed limit violations on this road type. However, it should be noted that although the European countries are supposed to collect data in a uniform manner, data collection procedures still vary substantially. Countries observe speeds for different vehicle types (all traffic together, cars and vans only) and different criteria are used to identify the measurement locations and appropriate (uncongested) traffic conditions [European Transport Safety Council, 2010].

5.2.1.3 Protective systems

Different from the above two behavioral characteristics which influence both the occurrence and severity of crashes, protective systems take effect especially when crashes happen. In case of a crash, the use of various protective systems by road users has been believed to play a vital role in protecting the most vulnerable parts of the human body against injury and considerably increasing the likelihood of surviving in serious crashes. Availability and appropriate use of protective systems (such as seat belts, child restraints, and helmets) are therefore fundamental items in developing related SPIs. First of all, it is estimated that seat belts have saved more than one million people that would have died in a road crash if not belted, thus being the biggest life saver on the roads [European Transport Safety Council, 2010]. In Europe, the use of seat belts is mandatory and it has been proven to provide a strong protection against fatalities in road crashes according to various studies. Elvik & Vaa (2004), for instance, indicated that the use of seat belts reduces the probability of being killed by 40-50% for drivers and front-seat passengers and by 25% for passengers in the rear seats. Moreover, regarding the use of safety seats for children and infants, studies (e.g., World Health Organization, 2004) have shown that infant deaths in cars are reduced by 70% and for small children by 50%. In addition, motorcycle helmets have been shown to have a clear impact on reducing fatal and serious head injuries by between 20% and 45% [World Health Organization, 2004]. The same study has also shown that bicycle helmets diminish the risk of head and brain injuries by 63% to 88%. Wearing helmets reduces the probability of being injured by around 25% [Elvik & Vaa, 2004]. Some European countries have legislated mandatory helmet use, which has been effective in preventing, or reducing injury severity of two-wheeler riders (motorcyclists, moped riders, and cyclists). Consequently, seat belt wearing rates in front seats and in rear seats, the proportion of child restraints use, and the usage rates of helmets by two-wheelers are the ideal SPIs related to the risk factor of the protective systems. However, due to data unavailability, three less ideal SPIs, i.e., *the daytime seat belt wearing rate in front and rear seats of light vehicles (<3.5 tons), respectively*, and *the daytime usage rate of child restraints,* are selected for this study, and the indicator data (average value of 2006-2008) for the 28 European countries are presented in Table 5.3.

Table 5.3 Protective systems related SPIs

	Seat belts		Child restraint
	Daytime seat belt wearing rate in front seats of light vehicles	Daytime seat belt wearing rate in rear seats of light vehicles	Daytime usage rate of child restraints
	06-08	06-08	06-08
AT	88.33%	57.33%	82.00%
BE	78.00%	40.00%	N/A
BG	85.00%	3.00%	30.00%
CY	80.00%	15.00%	N/A
CZ	88.67%	56.00%	42.00%
DK	91.00%	73.33%	N/A
EE	85.93%	55.20%	83.00%
FI	89.00%	81.33%	N/A
FR	97.67%	82.00%	89.00%
DE	96.33%	90.00%	84.00%
EL	68.50%	23.00%	N/A
HU	71.00%	41.00%	N/A
IE	88.00%	75.00%	N/A
IT	68.00%	30.00%	N/A
LV	79.00%	26.50%	N/A
LT	59.50%	30.00%	N/A

LU	80.00%	60.00%	57.00%
NL	93.67%	73.00%	72.00%
NO	91.50%	85.00%	94.00%
PL	78.00%	48.33%	86.00%
PT	86.00%	47.00%	N/A
RO	65.00%	5.00%	N/A
SK	68.00%	39.00%	N/A
SI	85.33%	45.17%	N/A
ES	86.33%	71.33%	N/A
SE	95.00%	76.00%	95.00%
CH	86.67%	66.00%	85.00%
UK	92.00%	85.67%	93.00%

Source: European Transport Safety Council (2010) and Vis & Eksler (2008)

We can see from Table 5.3 that all the 28 European countries collect data on the seat belt wearing rate, both in front seats and rear seats, and the rate for front seats is always higher than that for rear seats. However, still more than half of these countries have no available data on the usage rate of child restraints.

5.2.2 Vehicle

Vehicles, designed and used to transport people or cargo, are another important risk factor to road crashes. To some extent, a vehicle is inevitably involved in any road crash, but safer vehicles own more potential to prevent the occurrence of crashes as well as injuries in the event of a crash. Over the past decade, both vehicle active and passive safety have improved considerably in Europe due to increased minimum standards laid down by EU type approval regulations and vehicle manufacturers' efforts to meet consumer demands for safer vehicles [European Transport Safety Council, 2009]. Active safety features, such as anti-lock braking systems, traction control, driving aid systems and audible warning devices, help the driver in avoiding a road crash, while passive safety features better protect persons involved in the event of a crash, like frontal and side impact protection, airbags, load restraint and crush zones [Land Transport Safety Authority, 2000]. As the vehicle fleet is continuously being renewed to higher safety standards, the age of the vehicle fleet, or passenger cars in particular[17], represents a proxy for improvements in automotive engineering designed to resist the effect of crashes. That is, new cars tend to have more safety and protection features, and as the car ages, vehicular damage will increase. One study (in World Health Organization, 2004) showed that occupants in cars produced before 1984 run approximately three times the injury risk of new cars. In the current study, three SPIs related to the age of the passenger cars are developed to assess the performance of the vehicle fleet in different European countries. They are *the percentage of new passenger cars (less than 6 years)*, *the percentage of old passenger cars (more than 10 years)*, and *the annual renewal rate of passenger cars*.

Moreover, the crashworthiness of a vehicle has been used in many developed countries to assess the passive safety performance, i.e., how the vehicle performs in a crash situation. Many countries in the EU have set out legislation for safety standards in motor vehicles, such as the European New Car Assessment Programme, i.e., EuroNCAP, where vehicle crash performance is evaluated by rating the vehicle models according to their safety level for occupant protection, child protection, and pedestrian protection, respectively. Research by Lie & Tingvall (2000) concluded that vehicles gaining a higher ranking during EuroNCAP tests produce approximately 30% less fatal and serious injuries than low ranked vehicles. In this respect, six SPIs are selected in this study indicating some technical scores of new passenger cars from EuroNCAP tests. They are *the percentage of new passenger cars awarded 5 stars for occupant protection*, *the average percentage occupant protection score for new cars*, *the percentage of new passenger cars awarded 3 stars for pedestrian protection*, *the* average *percentage pedestrian protection score for new cars*, *the percentage of new passenger cars awarded 4 stars for child protection*, and *the percentage of new passenger cars with seat belt reminder (SBR)*.

[17] The availability of a wide range of information about passenger cars (e.g., age, weight, and size), and the fact that they make up the biggest proportion of the fleet in Europe makes passenger cars the most logical starting point in an assessment of the overall performance of the fleet in Europe.

Table 5.4 Vehicle related SPIs

	Age distribution of the vehicle fleet			Composition of the vehicle fleet		Technical scores					
						Occupant protection		Pedestrian protection		Child protection	Seat belt reminder
	% of new passenger cars: Less than 6 years	% of old passenger cars: More than 10 years	Annual renewal rate of passenger cars	% of goods vehicles in the vehicle fleet	% of powered two-wheelers in the vehicle fleet	% of new passenger cars awarded 5 stars for occupant protection	Average percentage occupant protection score for new cars	% of new passenger cars awarded 3 stars for pedestrian protection	Average percentage pedestrian protection score for new cars	% of new passenger cars awarded 4 stars for child protection	% of new passenger cars with seat belt reminder
	06-08	06-08	06-08	06-08	06-08	2008	2008	2008	2008	2008	2008
AT	39.80%	29.52%	7.07%	7.04%	12.61%	52%	89.3%	21%	36.1%	47%	70%
BE	39.51%	28.58%	10.47%	11.27%	6.10%	57%	89.9%	18%	34.2%	46%	73%
BG	N/A	N/A	1.89%	10.71%	3.71%	34%	83.8%	16%	34.7%	35%	53%
CY	27.96%	44.47%	5.67%	20.72%	7.28%	24%	93.7%	15%	42.8%	20%	33%
CZ	17.06%	57.46%	4.01%	9.65%	15.06%	29%	85.9%	14%	39.2%	37%	56%
DK	39.9%	31.42%	7.5%	18.78%	6.98%	46%	87.2%	21%	37.8%	41%	74%
EE	21.12%	61.16%	4.98%	13.18%	2.32%	43%	90.0%	25%	37.5%	53%	71%
FI	28.67%	44.77%	5.29%	11.78%	11.18%	60%	92.3%	22%	38.9%	60%	76%
FR	33.84%	33.59%	6.54%	13.62%	6.7%	59%	89.6%	23%	36.1%	45%	76%
DE	34.57%	33.65%	7.86%	5.06%	11.35%	55%	90.4%	19%	34.2%	47%	72%
EL	N/A	N/A	5.68%	17.05%	17.6%	39%	86.3%	26%	37.8%	40%	64%
HU	31.83%	44.42%	5.69%	12.66%	3.75%	38%	86.8%	32%	40.3%	39%	62%
IE	42.38%	16.57%	9.14%	14.9%	1.63%	62%	92.5%	23%	38.6%	56%	77%
IT	52.5%	22.97%	6.52%	8.96%	18.73%	47%	83.3%	19%	35.3%	29%	63%
LV	11.38%	76%	2.95%	11.89%	4.12%	40%	89.7%	19%	36.7%	48%	64%
LT	5.13%	85.9%	1.19%	7.97%	1.95%	43%	88.9%	20%	36.7%	52%	65%
LU	58.39%	16.19%	16.01%	8.23%	9.99%	59%	91.3%	18%	33.3%	48%	73%
NL	31.89%	33.77%	6.72%	10.33%	14.06%	52%	88.2%	23%	37.2%	45%	75%
NO	27.75%	44.19%	5.42%	17.17%	9.53%	62%	93.6%	23%	39.4%	61%	81%
PL	11.71%	65.83%	1.93%	13.91%	7.34%	48%	88.4%	23%	38.3%	45%	69%
PT	N/A	N/A	4.66%	21.33%	8.77%	59%	90.8%	28%	36.7%	51%	77%
RO	38.27%	41.84%	7.59%	12.69%	2.43%	25%	75.1%	13%	29.4%	30%	36%
SK	N/A	N/A	4.38%	12.56%	3.7%	32%	85.4%	23%	40.3%	35%	56%
SI	30.33%	34.17%	6.57%	6.64%	5.9%	52%	89.2%	24%	36.1%	45%	68%
ES	42.14%	28.74%	6.83%	16.69%	14.84%	58%	90.7%	27%	37.8%	51%	71%
SE	35.97%	32.73%	6.62%	9.42%	9.96%	64%	92.0%	14%	36.9%	56%	78%
CH	33.35%	32.83%	7.11%	6.51%	12.58%	49%	89.3%	19%	35.6%	42%	69%
UK	49.95%	17.98%	7.91%	11.11%	3.77%	54%	89.0%	22%	35.3%	46%	72%

Source: UNECE, 2011; European Commission, 2011a; and European Transport Safety Council, 2009

Another factor which influences the safety of the fleet is the proportions of vehicles of different types and weights that make up the total fleet, i.e., the vehicle composition [Vis, 2005]. The composition of the vehicle fleet also gives a rough indication of risk exposure on the road. It could be said that the relative share of motorcycles in use in the fleet is an indicator of the proportion of 'weak' motorized road users since motorcyclists are on average at greater risk of serious crashes, while the share of heavy goods vehicles, light goods vehicles, or other sport utility vehicles could be an aggressiveness indicator towards other road users [Hakkert et al., 2007a]. In this study, *the percentages of goods vehicles and powered two-wheelers in the vehicle fleet* are the two SPIs selected so as to reflect the composition of the vehicle fleet in each country.

Totally, 11 vehicle related SPIs are developed for the aforementioned three aspects, i.e., the age distribution of the vehicle fleet, the composition of the vehicle fleet, and some technical scores from EuroNCAP tests. They are presented in Table 5.4, along with the indicator data for the 28 European countries collected from the United Nations Economic Commission for Europe (2011), the European Commission (2011a), and the European Transport Safety Council (2009).

As can be seen, most of the countries have data on all these 11 vehicle indicators. However, for the indicators related to the technical scores from EuroNCAP tests, only the data for 2008 are available.

5.2.3 Infrastructure

The infrastructure of the road transport system, generally defined as the basic physical and organizational structures, or the facilities and services essential to enable, sustain, and enhance the daily road transport activities, is believed to have a strong impact on road safety as well. In this respect, road design and layout provide technical structures that make road transport possible, and influence crash risk as they determine how road users perceive the environment and offer instructions by means of signals [World Health Organization, 2004]. On the other hand, the level of available medical facilities and effective and timely emergency medical services are also a necessary infrastructure for sustainable road transport, which avoid preventable death and disability, and reduce the severity and suffering caused by the injury.

5.2.3.1 Road

The safety performance of the road transport system is the result of the combination of the functionality of the road network, homogeneity, and predictability of the road environment and the traffic involved [Hakkert et al., 2007a]. Four influencing factors are safety awareness in the planning of new road networks, dealing with safety features in the design of new roads, safety ameliorations to existing roads and healing actions on locations with a high accident risk. However, knowledge about the quantitative relations between road network, road design elements and road safety is still growing and by far not complete.

The road network consists of several road types. Motorways are despite their high speed limit considered to be the most safe type of roads. Fewer crashes resulting in fewer injuries happen on motorways than on other types of roads because of the separation between vehicle movements according to their speed. Elvik & Vaa (2004) showed that the rate of injury crashes per million vehicle kilometres of travel on motorways is about 25% of the average for all the public roads. However, they generally represent only a few percentages of the total road network. Rural roads account for a considerable share of all road fatalities. The risk of being killed (per kilometre driven) is generally higher on rural roads than on urban roads and is 4 to 6 times higher than on motorways [Organization for Economic Co-operation and Development, 2002].

In addition, poor road surface conditions as well as defects in road design and maintenance contribute to an increase in crash risk [European Transport Safety Council, 2001]. Objects along the road provide a risk in case the road user gets involved in a (run-of-the-road) crash. Bester (2001) reported that paved roads lead to lower fatality rates. Besides, some studies have assessed the safety performance of similar roads between countries by producing some sort of map or star rating for roads, such as the European Road Assessment Programme, i.e., EuroRAP, which aims at understanding the degree to which roads protect against severe injury in case of a crash [Lynam et al., 2004].

Table 5.5 Road related SPIs

	Motorway density (km/1000km^2)	Share of motorways and national roads in total road length
	06-08	06-08
AT	20.15	11.22%
BE	57.75	9.38%
BG	3.69	3.33%
CY	27.78	30.48%
CZ	8.37	5.28%
DK	25.30	4.28%
EE	2.20	5.51%
FI	2.11	17.03%
FR	20.13	2.04%
DE	35.26	8.23%
EL	8.24	9.58%
HU	10.45	4.04%
IE	4.56	5.63%
IT	21.87	5.41%
LV	0.00	2.35%
LT	4.73	6.37%
LU	56.84	34.15%
NL	62.80	3.76%
NO	0.79	28.77%
PL	2.14	7.06%
PT	28.16	11.05%
RO	1.10	20.08%
SK	7.32	8.56%
SI	30.48	4.06%
ES	25.43	3.85%
SE	4.00	4.78%
CH	33.32	2.47%
UK	14.90	12.5%

Source: European Commission, 2011a and European Union Road Federation, 2010

In this study, due to data unavailability, two SPIs related to motorways are used to represent the road performance of a country, which are *motorway density*, and *the share of motorways and national roads in total road length*. Indicator data (the average value of 2006-2008) for the 28 European countries are collected from the European Commission (2011a) and the European Union Road Federation (2010), as shown in Table 5.5.

Although all the countries have data on these two indicators, we should still keep in mind that they do not fully reflect the road performance of each country. More indicators related to road network, road design and maintenance, which are suitable for comparison among a large set of countries, have to be developed and refined in the future.

5.2.3.2 Emergency medical services

To make progress in road safety, apart from reducing risk before crashes (e.g., speed limit abidance) and during crashes (e.g., use of seat belts), improving the trauma management of people after crashes is also essential. In this respect, emergency medical services (EMS), concerned with the pre-hospital medical treatment of injuries resulting from road crashes, is recognized as a key component in avoiding preventable death and disability, and reducing the severity and suffering caused by the injury. A review of European case studies [European Transport Safety Council, 1999] concluded that about 50% of road traffic deaths occurred within a few minutes either at the scene of the crash or on the way to a hospital, 15% at the hospital within four hours of the crash and 35% after four hours. It means that many of these deaths might have been prevented if more immediate and better medical care would have been available. Studies worldwide [Hussain & Redmond, 1994; Mock et al., 1997] have shown that within the time period reaching a hospital, deaths and complications resulting in disability could be prevented in many cases. The European Commission (2003) has stated that several thousands of lives could be saved in the EU by improving the response times of the emergency services and other elements of post-impact care in the event of road traffic crashes. A review of 1970-1996 data in several OECD countries suggested that between 5% and 25% of the reductions in road crash deaths may have been due to improvements in medical care and technologies [Noland, 2004].

In this study, eight SPIs related to available medical facilities, professionally trained medical staff, suitable medical equipment, and timely emergency services are selected from the European *SafetyNet* project [Gitelman et al., 2008] characterizing the EMS performance of a country. They are *the number of EMS stations per 1000 km²*, *the percentage of EMS stations with at least one physician*, *the number of EMS medical staff per 10.000 citizens*, *the percentage of physicians and paramedics*, *the number of EMS transportation units per 100 km of road length*, *the percentage of high-equipped transportation units*, *the average response time*, and *the percentage of EMS response meeting the demand*. Indicator data, which were collected by means of a series of questionnaires distributed to the countries [Gitelman et al., 2008], are presented in Table 5.6.

Table 5.6 EMS related SPIs

	EMS stations		EMS staff		EMS transportation units		EMS response time	
	EMS stations per 1000 km^2	% of EMS stations with at least one physician	EMS medical staff per 10.000 citizens	% of physicians and paramedics	EMS transportation units per 100 km of road length	% of high-equipped transportation units*	average response time (min)	% of EMS response meeting the demand
	2006	2006	2006	2006	2006	2006	2006	2006
AT	5.0549	29%	51.316	29.4%	2.3010	100%	N/A	95%
BE	4.9463	26.5%	8.895	15%	0.3015	100%	6	100%
BG	1.9819	10%	9.553	22.4%	0.5450	86.6%	15	N/A
CY	1.9459	10%	4.185	19%	1.0391	100%	N/A	100%
CZ	2.4218	N/A	3.595	15.1%	0.4879	100%	7.83	89.2%
DK	3.2484	N/A	3.600	5.6%	0.6481	100%	8	100%
EE	1.1719	54.7%	9.935	18.4%	0.1599	100%	23	64%
FI	0.7387	2.4%	1.049	28.1%	0.5193	100%	N/A	N/A
FR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
DE	5.1302	39.4%	6.429	73.6%	1.1789	85.2%	8.10	91.5%
EL	0.0909	N/A	1.928	N/A	0.6499	99.1%	15	N/A
HU	2.3218	N/A	0.961	13.1%	0.5152	100%	16	72%
IE	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IT	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
LV	0.6506	10%	7.342	17.2%	0.3459	10%	17	88%
LT	0.9342	10%	4.746	18.8%	0.5293	10%	N/A	N/A
LU	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

NL	1.2281	N/A	1.620	N/A	0.4846	100%	N/A	N/A
NO	0.6177	N/A	N/A	N/A	0.7232	92.7%	N/A	90%
PL	0.6748	10%	2.104	N/A	0.6946	100%	N/A	90%
PT	5.2123	N/A	N/A	N/A	N/A	N/A	N/A	N/A
RO	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SK	6.9542	10%	7.089	21.4%	0.8590	53.2%	N/A	N/A
SI	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ES	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
SE	0.6107	N/A	4.407	0.2%	0.1200	100%	12.55	90%
CH	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UK	4.0153	N/A	4.650	64.2%	0.1512	21.4%	N/A	100%

*The high-equipped transportation units include Basic Life Support Units, Mobile Intensive Care Units and helicopters/planes.

Source: Gitelman et al. (2008)

Since only 21 European countries were considered in Gitelman et al. (2008), indicator data for the remaining seven countries (i.e., France, Ireland, Italy, Luxembourg, Romania, Slovenia, and Switzerland) are totally unavailable. Even for these 21 countries, only four countries (i.e., Belgium, Estonia, Germany, and Latvia) have provided complete data on these eight EMS indicators (only for the year 2006). With respect to the other countries, various levels of missing data exist, and the highest level is observed for Portugal, for which only one indicator value was provided. In addition, it should be noted that data collected from questionnaire surveys may produce considerable response bias due to for instance different definitions about some concepts in different countries. Data examination is therefore essential to determining their rationality and usefulness.

5.2.4 Summary

To develop a comprehensive set of safety performance indicators for international road safety programme benchmarking, three main components of the road transport system, i.e., road user behavior, vehicle and infrastructure, are considered, and six risk factors on the basis of these three components are identified, which are alcohol, speed, protective systems, vehicle, road, and emergency medical services. Several SPIs are then developed representing the characteristics of each of these six risk factors. Totally, 32 quantitative indicators are specified and they constitute a multilayer hierarchical structure as presented in Figure 5.2. In addition to the formulation of SPIs, available indicator data are collected (or calculated) for 28 European countries from a wide range of international data sources. In the following sections, some necessary data processing procedures are conducted before the road safety performance index can be constructed.

5.3 Data Processing

Given the high number of risk factors and corresponding SPIs, a large data set has to be collected, and a number of data errors are to be expected in spite of careful study design, conduct, and error-prevention strategies. In this study, two data processing procedures, i.e., statistical outlier detection and missing data treatment, are performed with the intention to identify and correct these errors or at least minimize their impact on the result of the road safety index research. After a full investigation, all instances with outliers (i.e., values or attributes that are far different from the expected values) are removed from the analysis, and missing values in the data set are imputed by using multiple imputation.

5.3.1 Outlier detection

In any data analysis, one of the first steps is the detection of outlaying observations, or *outliers*. Hawkins (1980) defines an outlier as *an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism*. Often, detected outliers are candidates for aberrant data that may otherwise adversely lead to model misspecification, biased parameter estimation and incorrect results. It is therefore important to identify them prior to modeling and analysis [Williams et al., 2002; Liu et al., 2004]. Over the last decades, a large number of outlier detection methods have been proposed [Hawkins, 1980; Rousseeuw & Leory, 1987; Caussinus & Roiz, 1990; Hadi, 1992; Barnett & Lewis, 1994]. One fundamental taxonomy is between *univariate methods* and *multivariate methods* [Ben-Gal, 2005]. Both of them are briefly introduced in the following sections, along with the applications for this study.

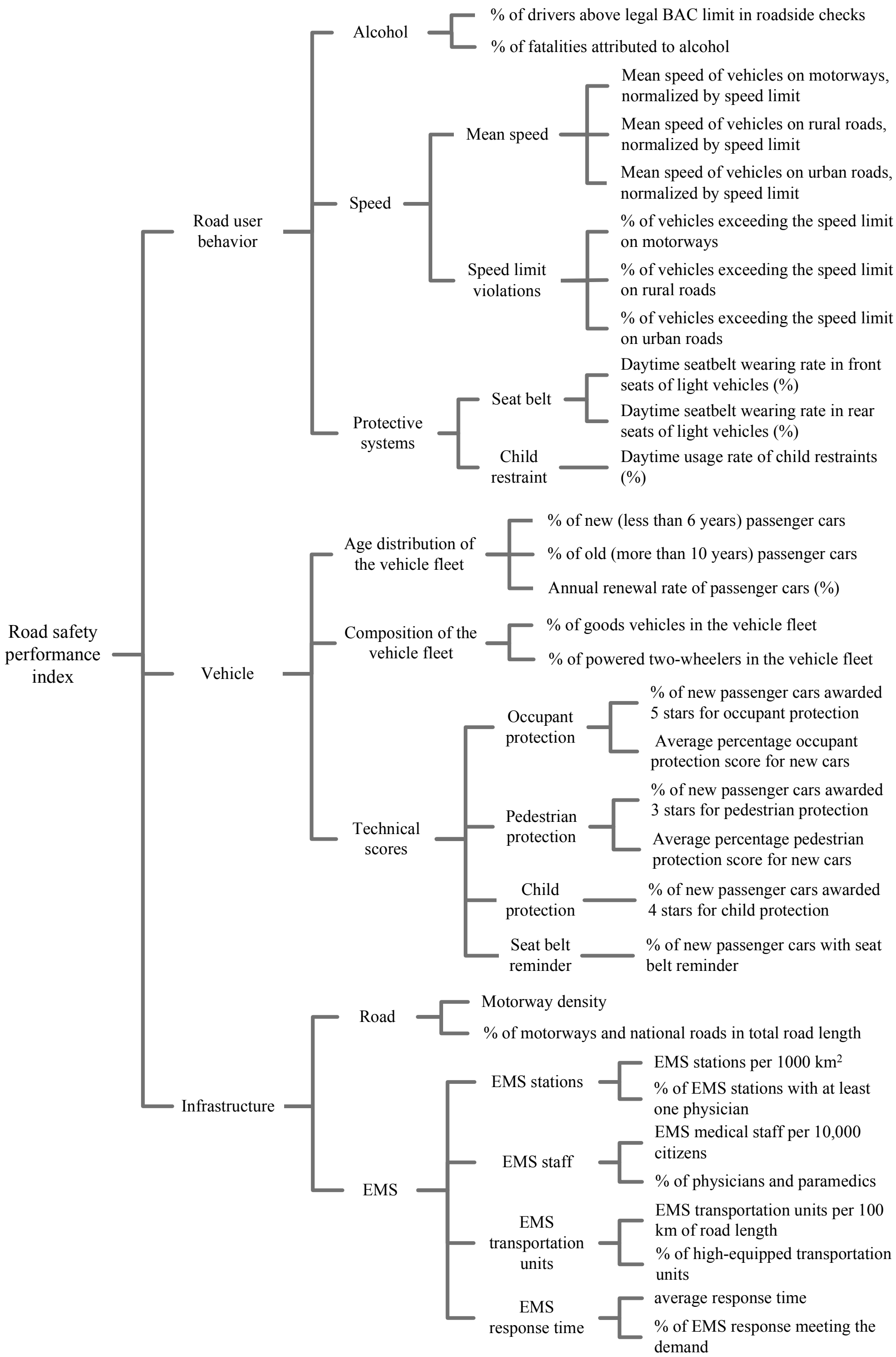


Figure 5.2 Hierarchical framework of safety performance indicators

5.3.1.1 Univariate methods

Most univariate analyses for outlier detection rely on the assumption of an underlying known distribution of the data, which is assumed to be identically and independently distributed. A central assumption in statistical-based univariate methods for outlier detection is a generating model that allows a small number of observations to be randomly sampled from distributions $G_1, \cdots, G_k$, differing from the target distribution *F*, which is often taken to be a normal distribution $N(\mu, \sigma^2)$ [Barnett & Lewis, 1994]. The outlier detection problem is then translated to the problem of identifying those observations that lie in a so-called outlier region. This leads to the following definition [Davies & Gather, 1993]: For any confidence coefficient α, $0 < \alpha < 1$, the α-outlier region of the $N(\mu, \sigma^2)$ distribution is defined by

$$outlier(\alpha, \mu, \sigma^2) = \{x : |x - \mu| > z_{1-\alpha/2}\sigma\} \quad (5\text{-}1)$$

where z_q is the *q* quintile of the *N*(0,1). A number *x* is an α-outlier with respect to *F* if $x \in outlier(\alpha, \mu, \sigma^2)$.

Based on this principle, one simple way to identify *univariate outliers* in practice is to convert all of the values for a variable *i* to *standard scores* (or *z*-scores) as follows:

$$z_i = \frac{x_i - \bar{x}}{sd} \quad (5\text{-}2)$$

where $\bar{x}$ denotes the sample mean, and *sd* the sample standard deviation. A case is generally considered as an outlier if its *z*-score exceeds 3.0 in absolute value [Schiffler, 1988].

To detect univariate outliers for this study, each of the 32 SPIs is standardized using Eq. (5-2). We thus obtain the *z*-scores of each country's indicator values. Five univariate outliers related to vehicle and EMS indicators are detected for four different countries with an absolute *z*-score higher than 3.0 (see Table 5.7).

Table 5.7 Five univariate outliers

Country	Indicator	z-score
AT	EMS medical staff per 10,000 citizens	3.88
AT	EMS transportation units per 100 km of road length	3.39
LU	Annual renewal rate of passenger cars	3.41
RO	Average percentage occupant protection score for new cars	-3.61
UK	% of high-equipped transportation units	-3.45

By checking the raw data set in Table 5-4 and Table 5-7, these indicator values are either extremely higher or greatly lower than others. Taking the EMS medical staff per 10,000 citizens in Austria as an example, a relatively high indicator value (51.316) was responded by Austria in the questionnaire survey conducted within the European *SafetyNet* project, while the average value of this indicator responded by all the other countries was only 4.829. It could probably happen due to a different definition of EMS medical staff in Austria. Consequently, all these five indicator values are omitted and treated as missing values in the following analysis.

5.3.1.2 Multivariate methods

In many cases multivariable observations cannot be detected as outliers when each variable is considered independently. Specifically, for a number of variables, the value for any of the individual variables may not be a univariate outlier, but, in combination with other variables, is a case that occurs very rarely. It is therefore called a *multivariate outlier*. Detection of such kind of outliers is only possible when multivariate analysis is performed, and the interactions among different variables are compared within the class of data.

As traditional multivariate outlier detection procedures, statistical methods often indicate those observations that are located relatively far from the center of the data distribution. Several distance measures can be implemented for such a task. The *Mahalanobis* distance is a well-known criterion which indicates the distance between a set of scores for an individual observation and the sample means for all variables. It is used as a diagnostic to assess for multivariate nonnormality

[Ben-Gal, 2005]. Given *n* observations from a *p*-dimensional data set, denote the sample mean vector by $\bar{x}_n$ and the sample covariance matrix by $\mathbf{V}_n$, where

$$\mathbf{V}_n = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)(x_i - \bar{x}_n)^T \tag{5-3}$$

The squared *Mahalanobis* distance for each multivariate data point *i*, *i*=1, …, *n*, denoted by D_i^2, is given as below:

$$D_i^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^T \mathbf{V}_n^{-1}(x_i - \bar{x}_n) \tag{5-4}$$

Since D_i^2 follows a chi-square distribution with degrees of freedom equal to the number of variables included in the calculation, a case would likely be a multivariate outlier if a significant D_i^2 score is obtained (at the $p<0.001$ level) [Peat & Bartion, 2005].

In this study, multivariate analysis is also performed to the 32 SPIs. Rather than examining each individual indicator, here, the SPIs in each category, such as the two alcohol indicators, the three indicators related to the mean speed, and so on, are considered simultaneously, and the probability for the *Mahalanobis* D_i^2 of these grouped indicators for each country is calculated in SPSS 17.0. We find that Cyprus has an unusual combination of its six indicator values corresponding to the technical scores of new passenger cars (with a probability of 0.0006), resulting in its designation as multivariate outlier. By checking the data source we used for these six indicators, i.e., the European Transport Safety Council (2009), Cyprus is the only country with a proportion of non-tested cars representing more than 50% of the new cars sold in 2008. As a result, these six indicator values for Cyprus are also excluded from our following data analysis.

5.3.2 Missing data treatment

To take as many of the available indicator information into account for road safety index research, a certain risk in the form of missing values is always present as no data collection system grants perfect data sets. Moreover, detecting and removing univariate and multivariate outliers from the original data set leads up to more missing values. The overall situation of missing data for this study is summarized in Figure 5.3.

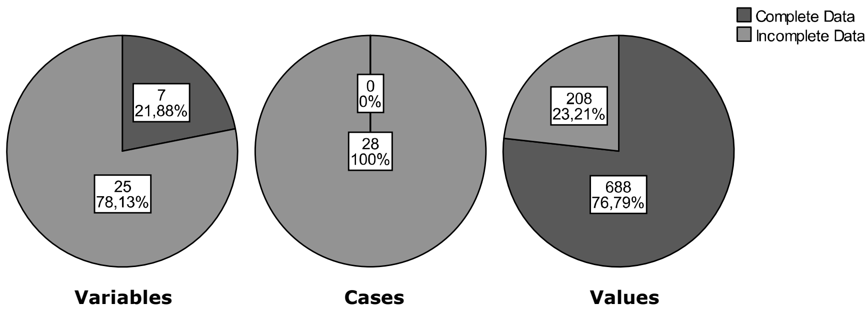


Figure 5.3 Overall summary of missing values

As can be seen, there are only 7 out of 32 SPIs for which all 28 countries have data, and no single country has values for the whole set of SPIs. Totally, around 23% of the indicator values are missing in the data set. Therefore, prior to the analysis, we have to decide whether to leave cases with missing data out of the analysis, or to replace the blank information by imputed values, as complete data matrices are in most cases the prerequisite of performing classical analyses.

During the last decades, various methods have been developed for handling missingness. The literature on the treatment of missing data is extensive and still in rapid development. Generally, three different strategies for dealing with missing data can be classified, which are *using as it is*, *deletion*, and *imputation* [Rodríguez et al., 2010; Kabak & Ruan, 2011].

More specifically, using the data as it is without any treatment is the most ideal strategy for missing data. In this way, original data sets with missing values are not preprocessed, i.e., data sets are not preliminarily converted into complete data sets. Thus, the models to apply on the data should be capable of using incomplete data, which however, is by no means easy and requires certain particular conditions. Since it is a premature approach for handling the general missing data issue, this strategy is still rare in literature and has been used only in limited application areas until now [Li, 2006; Grzymala-Busse, 2008; Kabak & Ruan, 2011].

On the contrary, a simple and common strategy for handling missingness is to delete cases containing any missing observations listwise or pairwise, and the analysis is then carried out on the data that remain [Little & Rubin, 1987; Schafer & Graham 2002; Howell, 2008; Oltman & Yahia, 2008]. These *ad hoc* methods, although simple to implement and being the default in the major statistical packages, have serious drawbacks in terms of elimination of useful information in the data and resulting in serious biases if the subjects who provide complete data are unrepresentative of the entire sample (i.e., the missing data are not missing completely at random [Rubin, 1976]). Consequently, as a rule of thumb, if a variable has more than 5% missing values, cases cannot be deleted [Little & Rubin, 1987]; many researchers are much more stringent than this. In such a case, rather than removing variables or observations with missing data, another strategy is to perform data imputation, defined as the process by which missing values in a data set are estimated by appropriately computed values, thereby producing a complete data set [Rubin, 1987]. Currently, most of the models dealing with missing data use this strategy. See also Chen & Shao (2000); Schafer & Graham (2002); Farhangfar et al. (2004); Molnar et al. (2008); Howell (2008); Jiang & Gruenwald (2008); Pospiech-Kurkowska (2008); Wang & Wang (2009); Silva-Ramírez et al. (2011), and their references.

Imputation has several desirable features. It is potentially more efficient than case deletion, because it uses 'expensive to collect' data that would otherwise be discarded, which helps to prevent loss of power resulting from a diminished sample size. Moreover, if the observed data contain useful information for predicting the missing values, an imputation procedure can make use of this information and maintain a high level of precision. Imputation also produces an apparently complete data set that may be analyzed by standard methods and software. To a data user, the practical value of being able to apply a favorite technique or software product can be immense. Finally, when data are to be analyzed by multiple persons or entities, imputing once, prior to all analyses, helps to ensure that the same set of units is being considered by each entity, facilitating the comparison of results. On the negative side, imputation can be difficult to implement well, particularly in multivariate settings. Some *ad hoc* imputation methods can distort data distributions and relationships. In the words of Dempster & Rubin (1983): "The idea of imputation is both seductive and dangerous. It is seductive because it can lull the user into the pleasurable state of believing that the data are complete after all, and it is dangerous because it lumps together situations where the problem is sufficiently minor that it can be legitimately handled in this way and situations where standard estimators applied to real and imputed data have substantial bias."

Over the past decades, a variety of imputation approaches have been proposed ranging from extremely simple to rather complex, such as unconditional mean imputation, regression imputation, hot deck imputation, decision trees imputation, clustering imputation, and neural networks imputation. All of them are known as *single imputation*, i.e., each missing value in a data set is replaced with one imputed value. If the simplicity is its main appeal, an important limitation of these methods is that subsequent analyses would fail to account for missing data uncertainty. Specifically, regardless of the single imputation method, imputed values are only estimates of the unknown true values. Any analysis that ignores the uncertainty of missing data prediction will lead to standard errors that are too small, *p*-values that are artificially low, and rates of Type I error that are higher than nominal levels [Schafer & Olsen, 1998]. To solve this problem, Rubin (1987) has developed the paradigm of *multiple imputation*. Instead of filling in a single value for each missing value, a multiple imputation procedure replaces each missing value with a list of simulated values that represent the uncertainty about the right value to impute (see Figure 5.4).

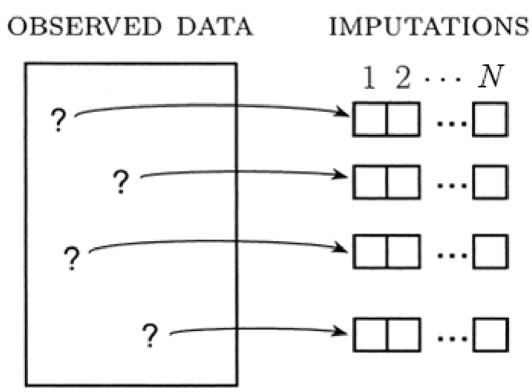


Figure 5.4 Schematic representation of multiple imputation

Source: Schafer & Olsen (1998)

Substituting the *j*th element of each list for the corresponding missing value, $j = 1, 2, \cdots, N$, produces *N* plausible alternative versions of the complete data. Each of the *N* data sets is then analyzed by using standard procedures for complete data and the results from these analyses are then combined. Multiple imputation retains much of the attractiveness of single imputation. However, it does not attempt to estimate each missing value through simulated values but rather to represent a random sample of the missing values. This process results in valid statistical inferences that properly reflect the uncertainty due to missing values, such as the valid confidence intervals for the parameters. Accordingly, multiple imputation is now becoming the dominant approach for the treatment of missing data. A further discussion of this method can be found in Rubin (1996), Schafer & Olsen (1998), Allison (2001), and Howell (2008).

In this study, we use the multiple imputation procedure in SPSS 17.0 [SPSS Inc., 2007] to impute all the missing values. Due to the relatively small number of countries with respect to the number of SPIs, the imputations are not straightforward but done separately for the indicators belonging to each of the three components of the road transport system, i.e., road user behavior, vehicle and infrastructure. The final data set that comprises both observed data and mean imputed values of the SPIs is presented in Table 5.8.

5.4 Conclusion

In this chapter, based on the identification of six leading road safety risk factors within the three road transport components, a comprehensive set of hierarchically structured safety performance indicators has been developed for capturing entire road safety risk in a country, and various international data sources providing indicator values for a large set of countries were consulted. Totally, 32 quantitative indicators were specified and available data collected (or calculated) for 28 European countries. Furthermore, outliers in the data set were examined by means of univariate and multivariate analyses, and missing values were imputed by using multiple imputation. The complete data set (see Table 5.8) provides us with the basis for the road safety index research discussed in the next chapters.

Table 5.8 Final data set for analysis, with original and mean imputed values of the SPIs

	Alcohol		Speed						Protective systems		
			Mean speed			% of speed limit violations			Seat belt		Child restraint
	% of drivers above the legal BAC limit in roadside checks	% of fatalities attributed to alcohol	on motorways	on rural roads	on urban roads	on motorways	on rural roads	on urban roads	Daytime seatbelt wearing rate in front seats	Daytime seatbelt wearing rate in rear seats	Daytime usage rate of child restraints
AT	0.074	0.077	0.910	0.896	1.031	0.213	0.194	0.539	0.883	0.573	0.820
BE	0.127	0.054	1.009	0.943	1.077	0.469	0.341	0.615	0.780	0.400	0.693
BG	0.070	0.042	0.932	0.977	0.615	0.720	0.504	0.572	0.850	0.030	0.300
CY	0.063	0.195	1.050	1.100	0.960	0.525	0.550	0.666	0.800	0.150	0.620
CZ	0.059	0.053	0.835	0.759	0.880	0.750	0.151	0.243	0.887	0.560	0.420
DK	0.043	0.248	0.936	1.056	1.039	0.315	0.698	0.600	0.910	0.733	0.726
EE	0.010	0.444	0.983	1.054	0.820	0.247	0.249	0.444	0.859	0.552	0.830
FI	0.015	0.260	0.886	0.960	0.912	0.399	0.439	0.423	0.890	0.813	0.680
FR	0.033	0.289	0.915	0.890	0.987	0.323	0.273	0.430	0.977	0.820	0.890
DE	0.093	0.116	0.943	0.836	1.062	0.588	0.388	0.555	0.963	0.900	0.840
EL	0.032	0.082	0.916	1.054	1.017	0.454	0.616	0.537	0.685	0.230	0.389
HU	0.031	0.126	0.858	0.883	1.012	0.451	0.301	0.594	0.710	0.410	0.411
IE	0.036	0.298	0.903	0.919	1.142	0.163	0.313	0.613	0.880	0.750	0.814
IT	0.088	0.036	1.012	0.990	1.074	0.375	0.444	0.357	0.680	0.300	0.684
LV	0.039	0.202	0.951	1.017	1.058	0.329	0.509	0.777	0.790	0.265	0.652
LT	0.016	0.111	0.854	0.982	1.158	0.207	0.394	0.430	0.595	0.300	0.384
LU	0.084	0.143	0.885	1.045	0.888	0.050	0.481	0.606	0.800	0.600	0.570
NL	0.106	0.036	0.950	0.945	0.939	0.360	0.491	0.583	0.937	0.730	0.720
NO	0.061	0.223	1.000	0.988	1.045	0.515	0.448	0.528	0.915	0.850	0.940
PL	0.095	0.081	1.059	1.004	1.278	0.563	0.658	0.826	0.780	0.483	0.860
PT	0.063	0.058	1.008	1.133	0.900	0.540	0.740	0.380	0.860	0.470	0.562
RO	0.122	0.084	0.941	0.982	0.924	0.285	0.401	0.576	0.650	0.050	0.446
SK	0.129	0.059	0.977	1.064	0.805	0.340	0.403	0.508	0.680	0.390	0.622
SI	0.070	0.456	0.885	0.700	1.160	0.340	0.010	0.840	0.853	0.452	0.638
ES	0.022	0.088	0.953	1.019	1.000	0.379	0.489	0.489	0.863	0.713	0.802
SE	0.009	0.100	0.966	0.977	0.950	0.678	0.526	0.528	0.950	0.760	0.950
CH	0.061	0.154	0.906	0.925	0.827	0.230	0.233	0.137	0.867	0.660	0.850
UK	0.169	0.156	0.994	0.797	1.006	0.517	0.100	0.493	0.920	0.857	0.930

	Vehicle										
	Age distribution of the vehicle fleet			Composition of the vehicle fleet		Technical scores					
						Occupant protection		Pedestrian protection		Child protection	SBR
	% of new passenger cars: Less than 6 years	% of old passenger cars: More than 10 years	Annual renewal rate of passenger cars	% of goods vehicles in the vehicle fleet	% of powered two-wheelers in the vehicle fleet	% of new passenger cars awarded 5 stars for occupant protection	Average percentage occupant protection score for new cars	% of new passenger cars awarded 3 stars for pedestrian protection	Average percentage pedestrian protection score for new cars	% of new passenger cars awarded 4 stars for child protection	% of new passenger cars with SBR
AT	0.398	0.295	0.071	0.070	0.126	0.520	0.893	0.210	0.361	0.470	0.700
BE	0.395	0.286	0.105	0.113	0.061	0.570	0.899	0.180	0.342	0.460	0.730
BG	0.203	0.583	0.019	0.107	0.037	0.340	0.838	0.160	0.347	0.350	0.530
CY	0.280	0.445	0.057	0.207	0.073	0.438	0.889	0.233	0.380	0.459	0.655
CZ	0.171	0.575	0.040	0.097	0.151	0.290	0.859	0.140	0.392	0.370	0.560
DK	0.399	0.314	0.075	0.188	0.070	0.460	0.872	0.210	0.378	0.410	0.740
EE	0.211	0.612	0.050	0.132	0.023	0.430	0.900	0.250	0.375	0.530	0.710
FI	0.287	0.448	0.053	0.118	0.112	0.600	0.923	0.220	0.389	0.600	0.760
FR	0.338	0.336	0.065	0.136	0.067	0.590	0.896	0.230	0.361	0.450	0.760
DE	0.346	0.337	0.079	0.051	0.114	0.550	0.904	0.190	0.342	0.470	0.720
EL	0.381	0.390	0.057	0.171	0.176	0.390	0.863	0.260	0.378	0.400	0.640
HU	0.318	0.444	0.057	0.127	0.038	0.380	0.868	0.320	0.403	0.390	0.620
IE	0.424	0.166	0.091	0.149	0.016	0.620	0.925	0.230	0.386	0.560	0.770
IT	0.525	0.230	0.065	0.090	0.187	0.470	0.833	0.190	0.353	0.290	0.630
LV	0.114	0.760	0.030	0.119	0.041	0.400	0.897	0.190	0.367	0.480	0.640
LT	0.051	0.859	0.012	0.080	0.020	0.430	0.889	0.200	0.367	0.520	0.650
LU	0.584	0.162	0.107	0.082	0.100	0.590	0.913	0.180	0.333	0.480	0.730
NL	0.319	0.338	0.067	0.103	0.141	0.520	0.882	0.230	0.372	0.450	0.750
NO	0.278	0.442	0.054	0.172	0.095	0.620	0.936	0.230	0.394	0.610	0.810
PL	0.117	0.658	0.019	0.139	0.073	0.480	0.884	0.230	0.383	0.450	0.690
PT	0.279	0.544	0.047	0.213	0.088	0.590	0.908	0.280	0.367	0.510	0.770
RO	0.383	0.418	0.076	0.127	0.024	0.250	0.842	0.130	0.294	0.300	0.360
SK	0.259	0.466	0.044	0.126	0.037	0.320	0.854	0.230	0.403	0.350	0.560
SI	0.303	0.342	0.066	0.066	0.059	0.520	0.892	0.240	0.361	0.450	0.680
ES	0.421	0.287	0.068	0.167	0.148	0.580	0.907	0.270	0.378	0.510	0.710
SE	0.360	0.327	0.066	0.094	0.100	0.640	0.920	0.140	0.369	0.560	0.780
CH	0.334	0.328	0.071	0.065	0.126	0.490	0.893	0.190	0.356	0.420	0.690
UK	0.500	0.180	0.079	0.111	0.038	0.540	0.890	0.220	0.353	0.460	0.720

	Road		EMS							
			EMS stations		EMS staff		EMS transportation units		EMS response time	
	Motorway density	% of motorways and national roads in total road length	EMS stations per 1000 km^2	% of EMS stations with at least one physician	EMS medical staff per 10.000 citizens	% of physicians and paramedics	EMS transportation units per 100 km of road length	% of high-equipped transportation units	average response time	% of EMS response meeting the demand
AT	20.148	0.112	5.055	0.290	8.741	0.294	0.767	1.000	11.199	0.950
BE	57.750	0.094	4.946	0.265	8.895	0.150	0.302	1.000	6.000	1.000
BG	3.694	0.033	1.982	1.000	9.553	0.224	0.545	0.866	15.000	0.880
CY	27.784	0.305	1.946	1.000	4.185	0.190	1.039	1.000	15.598	1.000
CZ	8.373	0.053	2.422	0.547	3.595	0.151	0.488	1.000	7.830	0.892
DK	25.299	0.043	3.248	0.734	3.600	0.056	0.648	1.000	8.000	1.000
EE	2.204	0.055	1.172	0.547	9.935	0.184	0.160	1.000	23.000	0.640
FI	2.107	0.170	0.739	0.024	1.049	0.281	0.519	1.000	13.173	0.700
FR	20.125	0.020	5.240	0.455	8.454	0.410	0.998	0.793	14.743	0.696
DE	35.256	0.082	5.130	0.394	6.429	0.736	1.179	0.852	8.100	0.915
EL	8.240	0.096	0.091	0.673	1.928	0.291	0.650	0.991	15.000	0.726
HU	10.450	0.040	2.322	0.521	0.961	0.131	0.515	1.000	16.000	0.720
IE	4.563	0.056	4.184	0.389	6.466	0.363	0.955	0.779	11.754	0.770
IT	21.870	0.054	4.498	0.448	6.149	0.429	0.920	0.855	11.796	0.773
LV	0.000	0.024	0.651	1.000	7.342	0.172	0.346	1.000	17.000	0.880
LT	4.732	0.064	0.934	1.000	4.746	0.188	0.529	1.000	12.479	0.921
LU	56.845	0.342	4.858	0.529	7.772	0.643	1.092	0.849	14.804	0.823
NL	62.796	0.038	1.228	0.363	1.620	0.301	0.485	1.000	13.803	0.568
NO	0.786	0.288	0.618	0.641	2.611	0.314	0.723	0.927	12.695	0.900
PL	2.143	0.071	0.675	1.000	2.104	0.261	0.695	1.000	12.039	0.900
PT	28.165	0.111	5.212	0.481	4.117	0.405	1.170	0.892	11.177	0.789
RO	1.105	0.201	3.504	0.479	4.543	0.266	0.931	0.858	14.446	0.771
SK	7.321	0.086	6.954	1.000	7.089	0.214	0.859	0.532	12.369	0.739
SI	30.484	0.041	6.472	0.292	6.138	0.444	0.666	0.703	14.381	0.609
ES	25.430	0.039	6.140	0.427	6.580	0.410	0.948	0.781	12.374	0.818
SE	3.998	0.048	0.611	0.624	4.407	0.002	0.120	1.000	12.550	0.900
CH	33.321	0.025	6.807	0.440	12.605	0.690	1.048	0.743	11.979	0.718
UK	14.904	0.125	4.015	0.650	4.650	0.642	0.151	0.745	12.509	1.000

To conclude, it is important to note that the selection of appropriate safety performance indicators requires periodic revisions, and the search for additional and better indicator data is an ongoing process as well. At this moment, indicators developed for most of the risk factors (except road) are extensive and comprehensive based on our current knowledge. However, reliable and comparable indicator data, especially concerning alcohol, speed, and emergency medical services, are still lacking to some extent. With respect to the factor of road, only limited and proxy indicators and data are currently available for benchmarking purposes. Knowledge on the quantitative relations between the road network, road design elements and road safety therefore needs further exploration, and a variety of appropriate indicators corresponding to this aspect call for different kinds of development efforts relating to concepts, methodologies, and data collection procedures. In addition, other risk factors that have a strong relationship with road safety or a large contribution to road crashes or casualties, such as inattentive driving as a result of mobile technology, could also be incorporated in the future and corresponding indicators developed and refined.

Chapter 6 Construction of a Composite Index (I): Hierarchical Structure Assessment[18]

This chapter elaborates on the use of a DEA model for composite index construction, especially when the hierarchical structure of the indicators is taken into account. It therefore answers the sixth research question of this dissertation. The proposed multiple layer DEA model is applied to combine the hierarchical SPIs developed in the previous chapter into a composite road safety performance index. Useful insights are gained from benchmarking analyses, and valuable recommendations are given enabling policymakers to prioritize their actions to enhance the level of road safety.

6.1 Introduction

In Chapter 5, a comprehensive set of current available national safety performance indicators was developed based on the identification of various underlying risk factors in road safety. Knowledge on these indicators is valuable in understanding the processes that lead to crashes and injuries, identifying corresponding interventions, and monitoring the effectiveness of the safety actions that are taken. However, since a number of indicators are considered for a particular risk factor in this study, simple comparisons per indicator only show a small piece of the road safety picture, which can be misleading since different countries may operate in different circumstances with different focal points. Consequently, a composite road safety indicator (or index), which combines individual indicator values into one single score, is valuable to be computed for the sake of meaningful benchmarking [Al-Haji, 2007; Wegman et al., 2008; Hermans, 2009a; Gitelman et al., 2010].

[18] Related research has been published in: Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K., (2011). A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: A case study in road safety performance evaluation, *Expert systems with applications*, Vol. 38, No. 12, pp. 15262-15272.

During the last decades, a large number of composite indexes (CIs) have been developed by various national and international organizations including United Nations (UN), Organization for Economic Cooperation and Development (OECD), World Health Organization (WHO), World Bank, and European Commission (EC), and been applied in wide ranging fields such as economy, society, governance, security, environment, sustainable development, globalization and innovation [Saisana & Tarantola, 2002; Freudenberg, 2003; Munda, 2005; Organization for Economic Co-operation and Development, 2008; Singh et al., 2009]. The proliferation of this kind of indexes is a clear symptom of their political importance and operational relevance in performance evaluation, benchmarking, and decision making. However, creating a CI, technically, is a mathematical aggregation of a set of individual indicators that measure multi-dimensional concepts but usually have no common units of measurement [Organization for Economic Co-operation and Development, 2008]. Therefore, the underlying construction scheme of a CI plays an important role and to a great extent determines the usefulness and credibility of the created CI.

The progress of recent studies on the development of a composite road safety index includes both objective methods (e.g., principal component analysis, factor analysis, neural networks and rough set theory) and subjective methods (e.g., analytical hierarchy process, budget allocation, and the technique for order preference by similarity to ideal solution) (see also Al-Haji (2007), Hermans et al. (2008), Gitelman et al. (2010), Shen et al. (2010), and Bao et al. (2012)). A point in common among these methods is that they all assign the same indicator weights for all the countries under study. It indeed enables the comparison among countries on a common base. However, in such a way, we make no full use of country-specific characteristics. In other words, the importance level of each indicator in each country is ignored, which makes the examination of root causes of poor performance in each country difficult.

In this respect, data envelopment analysis, which is based on self appraisal, has lately received considerable attention in the construction of CIs. The attractive features of DEA, relative to the other methods in developing CIs are: First, it provides a new way of combining multiple indicators without resorting to a priori knowledge on their tradeoffs, i.e., weights. Moreover, each country obtains its own best possible indicator weights, and DEA assesses the relative performance of a particular DMU by taking the performance of all other countries into account, which is known as the 'benefit of the doubt' (BOD) approach [Cherchye et al., 2007a]. In this way, key problems on road safety can be identified for each country separately, and policy-makers could not complain about unfair weighting, because each country is put in its most favorable light, and any other weighting scheme would generate a lower composite score. In other words, if a country turns out to be underperforming based on the most favorable set of weights, its poor performance cannot be traced back to an inappropriate evaluation process. Due to the aforementioned strengths, the applicability of DEA in CI construction has been widely explored in a number of recent studies such as the environmental performance index [Färe et al., 2004], the human development index [Despotis, 2005], the macro-economic performance index [Ramanathan, 2006], the sustainable energy index [Zhou et al., 2007], the internal market index [Cherchye et al., 2007b], the technology achievement index [Cherchye et al., 2008], and also a road safety performance index [Hermans, 2009a].

However, as today's performance management becomes more and more complex, a structural weakness of the basic DEA model has also arisen in its applications to CI construction. Specifically, due to the ever increasing complexity of numerous performance evaluation problems, more and more potential indicators might be used to represent an evaluation activity in a more comprehensive way. These indicators might also belong to different categories and further be linked to one another constituting a multilayer hierarchical structure. Under these circumstances, simply treating all the indicators to be in the same layer as is the case in the basic DEA model obviously ignores the information on their hierarchical structure, and further leads up to weak discriminating power and unrealistic weight allocations. To this end, Meng et al. (2008) introduced a layered hierarchy in the DEA model, in which the weights among categories were determined using the DEA method while the weights within categories (or internal weights) were determined by the weighted sum approach. However, this is a nonlinear model if all the weights are deduced from the mathematical model. Thereafter, Kao (2008) developed its linear transformation by introducing some variable substitutions. Nevertheless, the literature mentioned above was only limited to situations with a two-layer hierarchy, which might not entirely satisfy the need of increasingly complex evaluation problems. Shen et al. (2011) thereby further proposed a generalized multiple layer DEA (MLDEA) model to completely reflect the layered hierarchy within the DEA framework by incorporating different types of possible weight restrictions for each category of each layer.

In this chapter, starting from a brief introduction of the basic DEA-based CI model in Section 6.2, we elaborate the extension of the model for hierarchical structure assessment in Section 6.3. In Section 6.4, we demonstrate the application of this multilayer model to combine the hierarchical SPIs into a composite road safety performance index. Model outputs and corresponding road safety enhancing recommendations are subsequently presented in Section 6.5. The chapter ends with conclusions in Section 6.6.

6.2 DEA-based CI model

As introduced in Section 1.5, basic DEA models apply linear programming techniques to measure the relative efficiency of a set of DMUs on the basis of multiple inputs and multiple outputs. Therefore, to use DEA for CI construction, i.e., combining a set of individual indicators into one overall index, it means that only inputs or outputs of the DMUs will be taken into account in the model. As noted by Adolphson et al. (1991), it is possible to adopt a broader perspective, in which DEA is also appropriate for comparing any set of homogeneous units on multiple dimensions. Based on this perspective, Melyn & Moesen (1991) firstly introduced DEA to the field of CIs and the technique was applied to evaluate macroeconomic performance. Mathematically, the DEA-based CI model can be realized by converting the primal DEA model (1-4) into the following constrained optimization problem, which is known as the CCR model with constant inputs.

$$\begin{aligned} CI_c &= \max \sum_{r=1}^{s} u_r y_{rc} \\ s.t. \quad & \sum_{r=1}^{s} u_r y_{rj} \leq 1, \quad j = 1, \cdots, n \\ & u_r \geq 0, \quad r = 1, \cdots, s \end{aligned} \tag{6-1}$$

The n DMUs (or countries) are now to be evaluated by combining s different outputs (or indicators), with higher values indicating better performance, while the inputs of each DMU in model (1-4) are all assigned with a value of unity. This linear model is run n times to identify the optimal index score for all countries by selecting their best possible indicator weights separately, and the best-performing ones are those with an index score of one, while the others are underperforming.

Correspondingly, using the duality in linear programming, the envelopment form of the DEA-based CI model can be deduced as follows[19].

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} \lambda_j \\ s.t. \quad & \sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rc}, \quad r = 1, \cdots, s \\ & \lambda_j \geq 0, \quad j = 1, \cdots, n \end{aligned} \tag{6-2}$$

6.3 Multiple Layer DEA-based CI model

As a powerful performance measurement technique, the basic DEA-based CI model has received significant attention in recent years due to its prominent advantages over other traditional methods as presented in Section 6.1. However, in this model, all the indicators are equally treated as they belong to the same layer. It is acceptable when a low number of indicators is considered. As the amount grows, especially when a layered hierarchy is established, the hierarchical information on the indicators cannot be ignored arbitrarily, whereas the basic DEA-based CI model seems out of its capability to take this information into account. Consequently, the development of a multiple layer DEA-based CI model (MLDEA-CI) is desirable, which will be elaborated in the following sections.

6.3.1 The primal MLDEA-CI model

Suppose that a set of n DMUs is to be evaluated in terms of s indicators with a K layered hierarchy, which is shown in Figure 6.1.

[19] It can also be realized by assigning all the inputs with a value of unity in model (1-5).

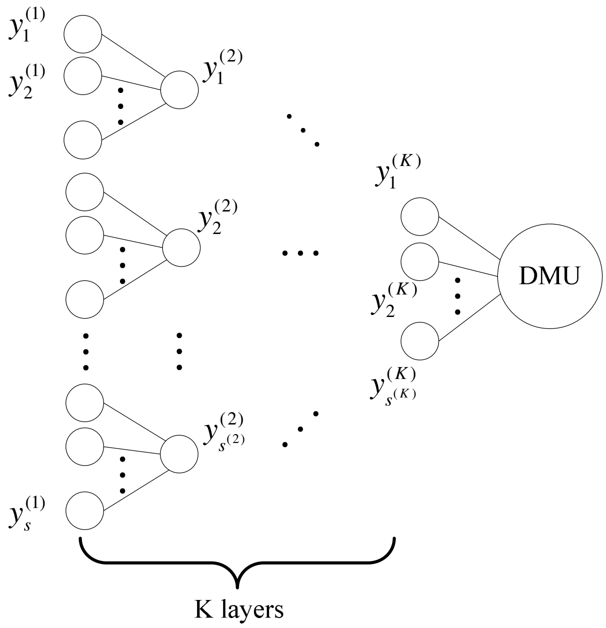


Figure 6.1 A hierarchical structure of indicators

where $s^{(k)}$ is the number of categories in the kth layer (k=1, 2, …, K), and $s^{(1)}=s$. The idea of the primal MLDEA-CI model is to first aggregate the values of the indicators within a particular category of a particular layer by the weighted sum approach in which the sum of the internal weights equals to one[20]. With respect to the final layer, the weights (or the multipliers) are determined using the basic DEA-based CI approach described in the previous section. Specifically, let $A_{f_k}^{(k)}$ denote the set of indicators of the fth category in the kth layer. The DMU_0's aggregated performance up to the Kth layer can then be expressed as:

$$
\begin{aligned}
& y_{f_K c}^{(K)} = \sum_{f_{K-1} \in A_{f_K}^{(K)}} p_{f_{K-1}}^{(K-1)} (\cdots \sum_{f_k \in A_{f_{k+1}}^{(k+1)}} p_{f_k}^{(k)} (\cdots \sum_{f_2 \in A_{f_3}^{(3)}} p_{f_2}^{(2)} (\sum_{f_1 \in A_{f_2}^{(2)}} p_{f_1}^{(1)} y_{f_1 c}^{(1)}))) \\
& \sum_{f_k \in A_{f_{k+1}}^{(k+1)}} p_{f_k}^{(k)} = 1, \quad p_{f_k}^{(k)} \geq 0, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1
\end{aligned} \tag{6-3}
$$

[20] The sum-up-to-one requirement for the internal weights is necessary for the following linear transformation. In doing so, normalized data should be used before aggregation so as to remove scale differences.

where $p_{f_k}^{(k)}$ denotes the internal weights associated with the indicators of the *f*th category in the *k*th layer, which are non-negative[21] and sum up to one within a particular category, so that each weight can be interpreted as the importance level of the corresponding indicator in the combined score.

Now, by substituting $y_{f_K c}^{(K)}$ from (6-3) to model (6-1), we obtain the following objective function:

$$CI_c = \max \sum_{f_K=1}^{s^{(K)}} u_{f_K} \Big(\sum_{f_{K-1} \in A_{f_K}^{(K)}} p_{f_{K-1}}^{(K-1)} \big(\cdots \sum_{f_k \in A_{f_{k+1}}^{(k+1)}} p_{f_k}^{(k)} \big(\cdots \sum_{f_2 \in A_{f_3}^{(3)}} p_{f_2}^{(2)} \big(\sum_{f_1 \in A_{f_2}^{(2)}} p_{f_1}^{(1)} y_{f_1 c} \big)\big)\big)\Big) \tag{6-4}$$

where u_{f_K} is the weight given to the *f*th category in the *K*th layer (i.e., the final layer), $f_k = 1, \cdots, s^{(k)}$.

To clearly illustrate the deduction process, we show in Figure 6.2 a simple example which contains eight indicators over three layers. Thus, K=3, $s^{(1)}$=8, $s^{(2)}$=4, and $s^{(3)}$=2.

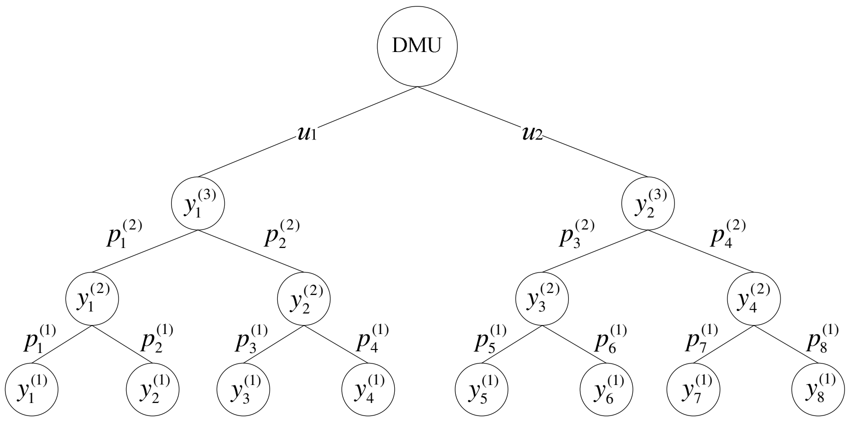


Figure 6.2 A DMU with three layers of indicators

We first calculate the aggregated values of the indicators in the last layer using (6-3) as: $y_1^{(3)} = p_1^{(2)} y_1^{(2)} + p_2^{(2)} y_2^{(2)} = p_1^{(2)}(p_1^{(1)} y_1^{(1)} + p_2^{(1)} y_2^{(1)}) + p_2^{(2)}(p_3^{(1)} y_3^{(1)} + p_4^{(1)} y_4^{(1)})$, and

[21] This condition can be replaced by using a small number $\zeta > 0$ for restricting the model to assign a weight of zero to unfavorable indicators.

$y_2^{(3)} = p_3^{(2)}y_3^{(2)} + p_4^{(2)}y_4^{(2)} = p_3^{(2)}(p_5^{(1)}y_5^{(1)} + p_6^{(1)}y_6^{(1)}) + p_4^{(2)}(p_7^{(1)}y_7^{(1)} + p_8^{(1)}y_8^{(1)})$, where $p_1^{(1)} + p_2^{(1)} = 1$, $p_3^{(1)} + p_4^{(1)} = 1$, $p_5^{(1)} + p_6^{(1)} = 1$, $p_7^{(1)} + p_8^{(1)} = 1$, $p_1^{(2)} + p_2^{(2)} = 1$, and $p_3^{(2)} + p_4^{(2)} = 1$.

Afterwards, the final index score of the DMU can be computed based on (6-4), which in this example is $u_1 y_1^{(3)} + u_2 y_2^{(3)}$, and it can be further expanded as: $u_1 p_1^{(2)} p_1^{(1)} y_1^{(1)} + u_1 p_1^{(2)} p_2^{(1)} y_2^{(1)} + u_1 p_2^{(2)} p_3^{(1)} y_3^{(1)} + u_1 p_2^{(2)} p_4^{(1)} y_4^{(1)} + u_2 p_3^{(2)} p_5^{(1)} y_5^{(1)} + u_2 p_3^{(2)} p_6^{(1)} y_6^{(1)} + u_2 p_4^{(2)} p_7^{(1)} y_7^{(1)} + u_2 p_4^{(2)} p_8^{(1)} y_8^{(1)}$.

However, since all the weights mentioned above are not given directly, their multiplication will lead up to a nonlinear model, and the more indicators to consider, the longer the iteration times and the harder to derive an optimal solution. To handle this problem, we introduce the following variable substitutions to linearize the model:

$$\hat{u}_{f_1} = \prod_{k=1}^{K-1} p_{f_k \in A_{f_{k+1}}^{(k+1)}}^{(k)} \cdot u_{f_K} \tag{6-5}$$

Summing up the weights of the indicators in each category of each layer (i.e., $p_{f_k}^{(k)}$), whose sum is equal to one, we obtain:

$$\begin{aligned} \sum_{f_1 \in A_{f_2}^{(2)}} \hat{u}_{f_1} &= \prod_{k=2}^{K-1} p_{f_k \in A_{f_{k+1}}^{(k+1)}}^{(k)} \cdot u_{f_K} \\ &\vdots \\ \sum_{f_1 \in A_{f_{K-1}}^{(K-1)}} \hat{u}_{f_1} &= p_{f_{K-1} \in A_{f_K}^{(K)}}^{(K-1)} u_{f_K} \\ \sum_{f_1 \in A_{f_K}^{(K)}} \hat{u}_{f_1} &= u_{f_K} \end{aligned} \tag{6-6}$$

Therefore, the weights of the indicators in each category of each output layer can be deduced as follows:

$$p_{f_k \in A_{f_{k+1}}^{(k+1)}}^{(k)} = \frac{\sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1}}{\sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1}}, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \tag{6-7}$$

To illustrate the above formulas, we still use the same example. According to (6-5), eight variable substitutions are needed in this case. They are: $\hat{u}_1 = u_1 p_1^{(2)} p_1^{(1)}$, $\hat{u}_2 = u_1 p_1^{(2)} p_2^{(1)}$, $\hat{u}_3 = u_1 p_2^{(2)} p_3^{(1)}$, $\hat{u}_4 = u_1 p_2^{(2)} p_4^{(1)}$, $\hat{u}_5 = u_2 p_3^{(2)} p_5^{(1)}$, $\hat{u}_6 = u_2 p_3^{(2)} p_6^{(1)}$, $\hat{u}_7 = u_2 p_4^{(2)} p_7^{(1)}$, $\hat{u}_8 = u_2 p_4^{(2)} p_8^{(1)}$. The index score of the DMU thus becomes $\sum_{i=1}^{8} \hat{u}_i y_i^{(1)}$. Moreover, since the sum of the internal weights in each category of each layer equals to one, i.e., $p_1^{(1)} + p_2^{(1)} = 1$, $p_3^{(1)} + p_4^{(1)} = 1$, $p_5^{(1)} + p_6^{(1)} = 1$, $p_7^{(1)} + p_8^{(1)} = 1$, $p_1^{(2)} + p_2^{(2)} = 1$, and $p_3^{(2)} + p_4^{(2)} = 1$, we can deduce that $\hat{u}_1 + \hat{u}_2 = u_1 p_1^{(2)}$, $\hat{u}_3 + \hat{u}_4 = u_1 p_2^{(2)}$, $\hat{u}_5 + \hat{u}_6 = u_2 p_3^{(2)}$, $\hat{u}_7 + \hat{u}_8 = u_2 p_4^{(2)}$, $\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4 = u_1$, and $\hat{u}_5 + \hat{u}_6 + \hat{u}_7 + \hat{u}_8 = u_2$ (see also (6-6)). Also, all the internal weights in each layer can now be computed by using (6-7). For instance, $p_1^{(2)} = (\hat{u}_1 + \hat{u}_2)/u_1 = (\hat{u}_1 + \hat{u}_2)/(\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4)$, and $p_1^{(1)} = \hat{u}_1/(u_1 p_1^{(2)}) = \hat{u}_1/(\hat{u}_1 + \hat{u}_2)$.

As indicated before, each weight assigned in a particular category of a layer can be interpreted as the importance level of the corresponding indicator. Therefore, the value judgment from decision makers or experts can be incorporated by restricting the weight flexibility in a particular category. In Section 4.3, some commonly used weight restriction techniques have been outlined, such as:

(i) Absolute weight restrictions, i.e., $L_{f_k}^{(k)} \le p_{f_k}^{(k)} \le U_{f_k}^{(k)}$, where $f_k \in A_{f_{k+1}}^{(k+1)}$, k=1, 2, …, K-1, L and U denote the lower respectively upper bounds of the weights;

(ii) Relative weight restrictions, i.e., $L_{f_k}^{(k)} \le p_{\alpha}^{(k)} / p_{\beta}^{(k)} \le U_{f_k}^{(k)}$, where $\alpha, \beta \in f_k \in A_{f_{k+1}}^{(k+1)}$, $\alpha \neq \beta$, k=1, 2, …, K-1, L and U are the lower and upper bounds, respectively; and

(iii) Ordinal weight restrictions, i.e., $p_{\alpha}^{(k)} \le \cdots \le p_{\beta}^{(k)}$, where, $\alpha, \beta \in f_k \in A_{f_{k+1}}^{(k+1)}$, $\alpha \neq \beta$, k=1, 2, …, K-1.

Now, by incorporating the deduced internal weights and appropriate weight restrictions into model (6-1), we obtain the primal MLDEA-CI model as follows:

$$
\begin{aligned}
& CI_c = \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c} \\
& s.t. \qquad \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j} \le 1, \quad j = 1, \cdots, n \\
& \qquad\qquad \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{p_{f_k}^{(k)}} \in \Theta, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \\
& \qquad\qquad \hat{u}_{f_1} \ge 0, \quad f_1 = 1, \cdots, s
\end{aligned}
\tag{6-8}
$$

This model reflects the layered hierarchy of the indicators by specifying the weights in each category of each layer. Meanwhile, by restricting the flexibility of these weights, denoted as Θ , consistency with prior knowledge and the obtainment of realistic and acceptable layer-specific weights are guaranteed, which cannot be realized in the one layer model. The model can be solved with a software package such as Lingo [Lindo Systems Inc., 2007].

In addition, based on (6-7), we can further prove that the three weight restriction formulas mentioned above will maintain the model to be linear. Specifically, the substitution of $p_{f_k}^{(k)}$ from (6-7) into constraint (i) results in the linear restriction below:

$$
L_{f_k}^{(k)} \cdot \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} \le \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \le U_{f_k}^{(k)} \cdot \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} \tag{6-9}
$$

where $f_k \in A_{f_{k+1}}^{(k+1)}$, k=1, 2, …, K-1.

Suppose the following absolute weight restriction in the previous example: $0.2 \le p_1^{(2)} \le 0.4$. It can then be substituted by: $0.2 \le (\hat{u}_1 + \hat{u}_2)/(\hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4) \le 0.4$, which can be further converted into the following two linear restrictions: $-0.8\hat{u}_1 - 0.8\hat{u}_2 + 0.2\hat{u}_3 + 0.2\hat{u}_4 \le 0$, and $0.6\hat{u}_1 + 0.6\hat{u}_2 - 0.4\hat{u}_3 - 0.4\hat{u}_4 \le 0$.

Likewise, substituting $p_{f_k}^{(k)}$ from (6-7) into constraint (ii) leads to:

$$
L_{f_k}^{(k)} \le \frac{\sum_{f_1 \in A_{\alpha}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1}}{\sum_{f_1 \in A_{\beta}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1}} \le U_{f_k}^{(k)} \tag{6-10}
$$

where $\alpha,\beta \in f_k \in A_{f_{k+1}}^{(k+1)}$, $\alpha \neq \beta$, k=1, 2, …, K-1. It can be noted that the denominators of $p_{\alpha}^{(k)}$ and $p_{\beta}^{(k)}$ are the same since they belong to the same category[22]. As a result, constraint (ii) can be written as:

$$L_{f_k}^{(k)} \cdot \sum_{f_1 \in A_{\beta}^{(k)}} \hat{u}_{f_1} \leq \sum_{f_1 \in A_{\alpha}^{(k)}} \hat{u}_{f_1} \leq U_{f_k}^{(k)} \cdot \sum_{f_1 \in A_{\beta}^{(k)}} \hat{u}_{f_1} \tag{6-11}$$

where $\alpha,\beta \in f_k \in A_{f_{k+1}}^{(k+1)}$, $\alpha \neq \beta$, k=1, 2, …, K-1.

Assume the following relative weight restriction for the same example: $1 \leq p_2^{(1)} / p_1^{(1)} \leq 2$. It equals to $1 \leq \frac{\hat{u}_2/(\hat{u}_1+\hat{u}_2)}{\hat{u}_1/(\hat{u}_1+\hat{u}_2)} = \hat{u}_2/\hat{u}_1 \leq 2$, and can be transformed into the linear constraint: $\hat{u}_1 \leq \hat{u}_2 \leq 2\hat{u}_1$, which reflects the situation of the ordinal constraint (iii) as well:

$$\sum_{f_1 \in A_{\alpha}^{(k)}} \hat{u}_{f_1} \leq \cdots \leq \sum_{f_1 \in A_{\beta}^{(k)}} \hat{u}_{f_1} \tag{6-12}$$

where $\alpha,\beta \in f_k \in A_{f_{k+1}}^{(k+1)}$, $\alpha \neq \beta$, k=1, 2, …, K-1.

6.3.2 The dual MLDEA-CI model

Having developed the primal MLDEA-CI model, we now further deduce its equivalent dual or envelopment form. In doing so, we firstly generalize the vector form of the primal model (6-8) as follows:

$$\begin{aligned} CI_c &= \max \hat{\boldsymbol{u}}\boldsymbol{y}_c \\ s.t. \quad & \hat{\boldsymbol{u}}\boldsymbol{Y} \leq 1 \\ & \hat{\boldsymbol{u}}\boldsymbol{Q} \leq 0 \\ & \hat{\boldsymbol{u}} \geq 0 \end{aligned} \tag{6-13}$$

where $\boldsymbol{Q}$ is a $s \times m$ weight restriction matrix corresponding to the second constraint in model (6-8), in which m denotes the total number of weight restrictions. In the aforementioned example,

[22] The weights of the factors are only comparable when they are in the same category.

$$
\boldsymbol{Q}=\begin{bmatrix} -0.8 & 0.6 & 1 & -2 & \cdots \\ -0.8 & 0.6 & -1 & 1 & \cdots \\ 0.2 & -0.4 & 0 & 0 & \cdots \\ 0.2 & -0.4 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \end{bmatrix}.
$$

Consequently, the dual MLDEA-CI model can be expressed as follows:

$$
\begin{aligned} \min \quad & \boldsymbol{e}_n\boldsymbol{\lambda} \\ s.t. \quad & \boldsymbol{Y\lambda}+\boldsymbol{Q\tau} \geq \boldsymbol{y}_c \\ & \boldsymbol{\lambda} \geq 0, \quad \boldsymbol{\tau} \geq 0 \end{aligned} \tag{6-14}
$$

where $\boldsymbol{e}_n$ is a row vector ($1\times n$) with all elements equal to one, $\boldsymbol{\lambda}=(\lambda_1,\lambda_2\cdots\lambda_n)^T$ is the dual weight vector with the same definition as in model (6-2), and $\boldsymbol{\tau}=(\tau_1,\tau_2,\cdots\tau_m)^T$ is an extra vector due to the incorporation of weight restrictions in model (6-13).

6.4 Application

In this study, we apply the proposed MLDEA-CI model to combine the 32 hierarchical SPIs developed in the previous chapter into a composite road safety performance index so that the overall road safety performance of the 28 European countries can be evaluated. In doing so, data normalization and weight restrictions are two necessary steps that need to be specified before the model can be applied.

6.4.1 Data normalization

Prior to the application of the MLDEA-CI model, the raw data should be first normalized so as to eliminate the scale differences of the indicators and the effects of the measurement unit, and moreover, to ensure that all the indicators are expressed in the same direction with respect to their expected road safety impact, i.e., a high performance indicator value should always correspond to a low crash/injury risk.

A large number of normalization methods have been proposed in literature such as rescaling, standardization, and ranking [Freudenberg, 2003]. In this study, the distance to a reference approach [Organization for Economic Co-operation and Development, 2008] is adopted because the ratio of two numbers is best kept by this approach:

$$\tilde{y}_{rj} = \begin{cases} y_{rj} / y_j^*, \forall_j, y_j \textit{ is a benefit indicator} \\ y_j^- / y_{rj}, \forall_j, y_j \textit{ is a } \text{cost} \textit{ indicator} \end{cases} \tag{6-15}$$

where $\tilde{y}_{rj}$ are normalized indicator values. y_j^* and y_j^- are the maximum and minimum values of each indicator in the data set, which are selected as the reference (or benchmark) for normalization when a benefit respectively a cost indicator is taken into account. Based on (6-15), the original data set presented in Table 5.8 can be normalized, in which all the eight indicators related to alcohol and speed, as well as three vehicle indicators (i.e., the proportion of old passenger cars, the proportion of goods vehicles, and the proportion of powered two-wheelers) and one EMS indicator (i.e., the average response time) are identified as cost indicators for this study, while the others are benefit indicators. As a result, the country with the highest safety performance receives a normalized value of one whereas the others are expressed as percentage share of that country's value. Taking the percentage of road fatalities attributed to alcohol as an example, the Netherlands performs best (1.000) while Slovenia worst (0.078), and all other countries' values lie within this interval.

6.4.2 Weight restrictions

In addition to the data normalization, weight restrictions for each layer of the indicators should be specified before applying the MLDEA-CI model so that the obtainment of realistic and acceptable indicator weights is guaranteed. In this study, the SPIs belonging to the same category of each layer (except the last layer) are considered to be of similar importance, such as the two indicators with respect to alcohol (i.e., the percentage of drivers above the legal BAC limit in roadside checks and the percentage of road fatalities attributed to alcohol), the two aspects related to speed (i.e., the mean speed and the speed limit violations), as well as the three risk factors of road user behavior (i.e., alcohol, speed, and protective systems). Thus, the absolute weight restrictions (i) introduced in Section 6.3.1 are applied, in which we obligate the weights to vary within a range from 0.8 to 1.2 of their average weights. Taking the mean speed as an example, the weights of its three indicators (the average weight is 0.333) are thereby required to lie between 0.267 and 0.4. With regard to the last layer, i.e., the three main components of the road transport system – road user behavior (*R*), vehicle (*V*), and infrastructure (*I*) – a combination of the relative weight restrictions (ii), the ordinal weight restrictions (iii) and the virtual weight (or share) restrictions introduced in Section 4.3.4 is assigned, i.e., $Share_R > 2Share_I > 4Share_V > 20\%$, indicating the importance ordering of these three components in the contribution of road crashes (see also Figure 5.1).

6.5 Results

By now, the proposed MLDEA-CI model (both the primal and dual form) can be applied to determine the most optimal road safety performance index score for each of the 28 European countries by taking all the 32 hierarchical SPIs into account. Best-performing countries are subsequently distinguished from underperforming ones and countries ranked. Moreover, useful benchmarks for the underperforming countries can be identified by using the dual model (6-14), and the indicator weights allocated in each layer of the hierarchy can be deduced for each country based on the primal model (6-8). All the model outputs are illustrated in the following sections, and further translated into road safety enhancing recommendations.

6.5.1 Index scores and country ranking

With the developed MLDEA-CI model, the 32 normalized indicator values are now combined into a composite index score for each country by selecting the best possible indicator weights under the imposed restrictions. The results are shown in Table 6.1, along with the ones from the basic DEA-CI model (6-1)[23].

[23] In this model, the same weight restrictions on three main components of the road transport system are imposed.

Table 6.1 Road safety performance index score of the 28 European countries based on the basic DEA-CI model and the MLDEA-CI model

	Index score based on the basic DEA-CI model	Index score based on the MLDEA-CI model
AT	1.000	0.965
BE	1.000	0.958
BG	1.000	0.869
CY	1.000	0.834
CZ	1.000	0.912
DK	1.000	0.884
EE	1.000	0.928
FI	1.000	0.906
FR	1.000	0.944
DE	1.000	1.000
EL	0.999	0.749
HU	1.000	0.751
IE	1.000	0.905
IT	1.000	0.967
LV	1.000	0.746
LT	1.000	0.805
LU	1.000	1.000
NL	1.000	1.000
NO	1.000	0.941
PL	1.000	0.845
PT	1.000	0.903
RO	0.986	0.766
SK	1.000	0.877
SI	1.000	0.913
ES	1.000	0.961
SE	1.000	1.000
CH	1.000	1.000
UK	1.000	0.971

Since a large number of indicators relative to the number of countries is considered in this study, most of the countries (except for Greece and Romania) obtain the index score of one based on the basic DEA-CI model, which implies its weak capability of discriminating between countries in terms of their road safety performance. By applying the MLDEA-CI model, however, due to the consideration of hierarchical information on these indicators and the incorporation of corresponding weight restrictions, the discriminating power of the model is obviously improved and the optimal index score of one is obtained by only five best-performing countries. They are Germany, Luxembourg, the Netherlands, Sweden, and Switzerland.

Having identified the best-performing and underperforming countries, we further rank all these countries by applying the cross efficiency matrix with the aggressive formulation introduced in Section 2.4.2. The cross-index scores of these 28 European countries are therefore computed to reflect their all round road safety performance by taking the best possible weights for each country in the data set into account, which are shown in the second column of Table 6.2 in decreasing order. Sweden obtains the highest score (0.996), and is thereby ranked at the top, while Latvia the worst (0.7). Although they both belong to Northern Europe based on the United Nations Statistics Division (http://unstats.un.org/unsd/methods/m49/m49regin.htm#europe), generally speaking, the Western European countries (e.g., the Netherlands and Germany) have a relatively higher index score than the Eastern European countries (e.g., Romania and Hungary), and the Northern European countries, especially those Nordic countries (e.g., Sweden and Norway) are better performing than the Southern European countries (e.g., Greece and Slovenia).

Table 6.2 Cross-index score and cross-efficiency score of the 28 European countries

	Cross-index score	Cross-efficiency score
SE	0.996	0.940
NL	0.984	0.985
LU	0.983	0.675
CH	0.973	0.950
DE	0.960	0.773
ES	0.931	0.503
AT	0.924	0.514
UK	0.911	0.930
BE	0.908	0.463
IT	0.904	0.577
FR	0.902	0.644
EE	0.885	0.897

NO	0.884	0.298
PT	0.864	0.455
IE	0.862	0.588
FI	0.858	0.704
CZ	0.851	0.357
DK	0.840	0.609
SI	0.836	0.372
SK	0.822	0.301
PL	0.791	0.262
CY	0.779	0.373
BG	0.776	0.249
LT	0.757	0.234
RO	0.709	0.218
HU	0.707	0.281
EL	0.702	0.286
LV	0.700	0.227

As a relevant point of reference, the overall road safety efficiency score proposed in Chapter 2, which considers the road safety final outcomes (e.g., the number of road fatalities) on the one hand, and the three common measures of exposure to risk, i.e., the number of inhabitants, passenger-kilometres travelled and passenger cars on the other hand, is recomputed here based on the average values of 2006-2008, and the cross-efficiency score of these 28 European countries is presented in the last column of Table 6.2. The high degree of consistency between these two sets of scores is verified by their significant correlation coefficient (0.806), which further implies that the created road safety performance index has a clear link with road safety output, and can be used as a valuable predictor based on which efficient policy measures can be put forward.

6.5.2 Relevant benchmarks

To better understand the computational process leading to the index scores presented in the last column of Table 6.1, especially the reasons why the 23 underperforming countries are unable to obtain a value of one, we further explore the mechanism of the MLDEA-CI model. Theoretically, the primal MLDEA-CI model (6-8) is used to determine the best possible indicator weights under the imposed restrictions that maximize the index score of a certain country. In doing so, the performance of all other countries is taken into account. Therefore, if the optimal weights of a country *A* under study do not result in a value of one for this country but cause the weighted score of another country *B* in the data set to become one, then the model stops. This implies that country *B* is characterized by higher road safety performance than country *A* with respect to at least one risk aspect since the index score of *B* is relatively higher with the same set of weights. Therefore, country *B* can be viewed as a realistic and valuable benchmark for country *A*, and a series of benchmark countries like *B* constitute a reference set for country *A* to learn from. Moreover, based on the dual weights (i.e., λ) calculated from the dual MLDEA-CI model (6-14), the relative importance of a benchmark country within the reference set can be identified. Taking Austria as an example, the best possible indicator weights assigned for this country only result in its optimal index score of 0.965 because a weighted score of one is achieved by four other countries, which are Luxembourg, the Netherlands, Sweden, and Switzerland. Therefore, Austria is underperforming, and it could take a hypothetical composite country which consists of the above four best-performing countries as an example. Among others, Sweden appears to be the most important benchmark as it obtains the greatest dual weight, which is 0.547. In other words, Austria should learn the most from Sweden to improve its road safety performance. Table 6.3 indicates the reference set for each of the 23 underperforming countries, and the benchmark country with the greatest dual weight is highlighted.

Table 6.3 Benchmarks for the underperforming countries

	DE	LU	NL	SE	CH
AT		0.280	0.009	**0.547**	0.129
BE		0.123	**0.711**	0.110	0.014
BG			**0.777**	0.092	
CY		0.067	0.043	**0.723**	
CZ			**0.711**	0.200	
DK		0.064	0.236	**0.584**	
EE				**0.928**	
FI		0.068	0.027	**0.810**	
FR	0.159	0.089	0.146	**0.551**	
EL			**0.400**	0.349	
HU		0.237	0.185	**0.329**	

IE	0.044	0.207		**0.654**	
IT			**0.942**	0.025	
LV		0.167		**0.579**	
LT			0.151	**0.654**	
NO				**0.941**	
PL			0.032	**0.813**	
PT		0.275	**0.483**	0.085	0.060
RO		0.139	**0.428**	0.127	0.072
SK		0.178	0.307	**0.365**	0.027
SI		**0.373**		0.221	0.319
ES		0.072	0.330	**0.508**	0.051
UK	0.253			**0.718**	

6.5.3 Weight allocation and road safety priorities

In addition to the identification of specific benchmarks for all the underperforming countries, we further explore the indicator weights allocated in each layer of the hierarchy per country from the view of the primal MLDEA-CI model (6-8). That way, areas of underperformance (also for those best-performing countries) can be detected, and road safety priorities for policy action can be formulated. More specifically, in the basic DEA-CI model (6-1), all indicators are simply treated to be in the same layer and no layer related weight restrictions can be imposed. Therefore, weights will be allocated with the only purpose of maximizing the index score regardless of the position of the indicators in the hierarchical structure. On the contrary, the MLDEA-CI model not only pursues the optimal index scores, but also guarantees its consistency with prior knowledge and the obtainment of realistic and acceptable weights by restricting the weight flexibility in each category of each layer. More importantly, based on the principle of the MLDEA-CI model, an indicator is assigned a high weight if the country performs relatively well on that aspect. On the contrary, low weights provide policymakers with valuable information about the aspects requiring most action for improvement. In Figure 6.3, the assigned weights (the values in brackets are shares) based on the primal MLDEA-CI model are presented for the case of Austria, which obtains the optimal index score of one in the basic DEA-CI model, while a lower value (0.965) in the multilayer model.

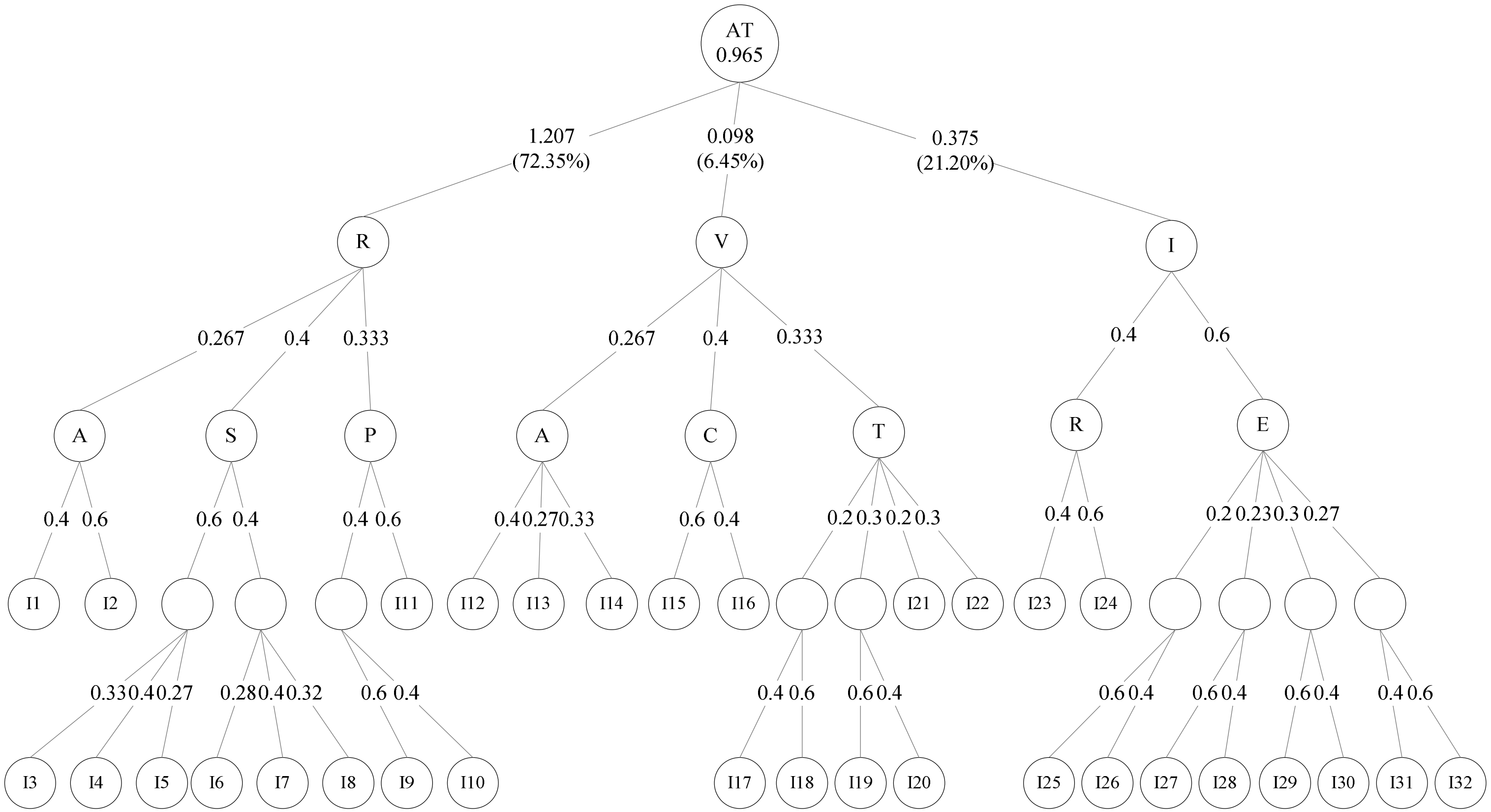


Figure 6.3 Assigned weights (and shares) in each layer of the hierarchy for Austria based on the MLDEA-CI model

Figure 6.3 shows the accordance of the weights (and shares) with the imposed restrictions described in Section 6.4.2. For instance, the indicators belonging to a particular category of each layer are of similar importance (with a 20% variability of their average weights), such as the three mean speed indicators (*I3*-*I5*). Also, the share of road user behavior (72%) is more than twice as large than that of infrastructure (21%), which is also more than two times greater than that of vehicle (6%). Moreover, since each weight allocated in a particular category of a layer in the MLDEA-CI model can be interpreted as the importance share of the corresponding indicator, more detailed insight can be gained based on these weights. Still taking the indicators *I3*-*I5* as an example, the assigned weights imply that *I5*, i.e., the mean speed on urban roads should be given priority over the other two indicators in terms of Austrian road safety policy action since the lowest weight (0.267) is allocated to this indicator. Considering all the 11 SPIs related to road user behavior by the same principle, we find that Austria is doing relatively well in the speed aspect (with the highest weight of 0.4), especially the mean speed on rural roads (*I4*). Whereas more policy attention should be paid to the risk aspect on alcohol (with the lowest weight of 0.267), followed by the protective systems (with a weight of 0.333), in which improving rear seat belt wearing rate (*I10*) is most urgently needed. Based on the same principle, road safety priorities with respect to the other two components (i.e., vehicle and infrastructure) in Austria can be identified as well. They are to reduce the proportion of old passenger cars in the vehicle fleet (*I13*) and to raise the motorway density (*I23*), respectively. To conclude, although a relatively lower index score is achieved by using the MLDEA-CI model, the multilayer model is to be preferred since it produces more valuable results by means of taking all the indicators and their hierarchical structure into account. In Table 6.4, the layer-specific road safety priorities for all the 28 European countries are summarized, in which 1-1 represents the most urgent road safety performance aspect with respect to the most important risk factor, 2-1 represents the most urgent road safety performance aspect with respect to the second most important risk factor, and so on.

Table 6.4 Road safety priorities in each of the 28 European countries

	Alcohol		Speed						Protective systems		
			Mean speed			% of speed limit violations			Seat belts		Child restraint
	% of drivers above legal BAC limit in roadside checks	% of fatalities attributed to alcohol	on motorways	on rural roads	on urban roads	on motorways	on rural roads	on urban roads	Daytime seatbelt wearing rate in front seats	Daytime seatbelt wearing rate in rear seats	Daytime usage rate of child restraints
AT	1-1				3-2-1	3-1-1				2-1-1	
BE	1-1				3-2-1			3-1-1		2-1-1	
BG	2-1			3-2-1		3-1-1					1-1
CY	1-1			3-2-1		3-1-1				2-1-1	
CZ	2-1				3-2-1	3-1-1					1-1
DK	1-1				2-2-1		2-1-1				3-1
EE		1-1		3-2-1			3-1-1			2-1-1	
FI	1-1			3-2-1			3-1-1				2-1
FR	1-1				2-2-1		2-1-1				3-1
DE	1-1				2-2-1			2-1-1			3-1
EL	2-1			3-2-1			3-1-1				1-1
HU	1-1				3-2-1	3-1-1					2-1
IE	1-1				2-2-1			2-1-1		3-1-1	
IT	2-1				3-1-1		3-2-1			1-1-1	
LV	1-1				3-2-1			3-1-1		2-1-1	
LT	2-1				3-2-1		3-1-1				1-1
LU		1-1		3-1-1				3-2-1			2-1
NL	3-1		1-1-1					1-2-1			2-1
NO	1-1				2-2-1		2-1-1		3-1-1		
PL	1-1				2-2-1			2-1-1		3-1-1	
PT	1-1			3-2-1		3-1-1					2-1
RO	1-1			3-2-1		3-1-1				2-1-1	
SK	1-1			3-2-1			3-1-1			2-1-1	
SI	1-1				3-1-1	3-2-1					2-1
ES	1-1			2-2-1				2-1-1	3-1-1		
SE		3-1	1-1-1					1-2-1	2-1-1		
CH		1-1		3-1-1			3-2-1		2-1-1		
UK	1-1				2-2-1			2-1-1	3-1-1		

	Vehicle										
	Age distribution of the vehicle fleet			Composition of the vehicle fleet		Technical scores					
						Occupant protection		Pedestrian protection		Child protection	SBR
	% of new passenger cars: Less than 6 years	% of old passenger cars: More than 10 years	Annual renewal rate of passenger cars	% of goods vehicles in vehicle fleet	% of powered two-wheelers in vehicle fleet	% of new passenger cars awarded 5 stars for occupant protection	Average percentage occupant protection score for new cars	% of new passenger cars awarded 3 stars for pedestrian protection	Average percentage pedestrian protection score for new cars	% of new passenger cars awarded 4 stars for child protection	% of new passenger cars with SBR
AT		1-1			3-1	2-1-1				2-1	
BE		3-1			2-1			1-1-1			1-1
BG		1-1			3-1					2-1	2-1
CY	1-1				2-1					3-1	3-1
CZ	1-1				3-1					2-1	2-1
DK		2-1		1-1		3-1-1				3-1	
EE		1-1		3-1		2-1-1					2-1
FI	1-1				2-1	3-1-1					3-1
FR	1-1			2-1						3-1	3-1
DE	1-1				3-1					2-1	2-1
EL		2-1			1-1					3-1	3-1
HU		1-1			2-1					3-1	3-1
IE	1-1			3-1						2-1	2-1
IT			2-1		3-1					1-1	1-1
LV	1-1			2-1		3-1-1					3-1
LT		1-1			3-1					2-1	2-1
LU	1-1				3-1					2-1	2-1
NL		1-1			2-1					3-1	3-1
NO	1-1				2-1					3-1	3-1
PL	1-1				2-1					3-1	3-1
PT		1-1		2-1						3-1	3-1
RO		1-1		3-1						2-1	2-1
SK			1-1	3-1						2-1	2-1
SI	1-1				3-1					2-1	2-1
ES			1-1	2-1						3-1	3-1
SE		2-1			1-1			3-1-1			3-1
CH	1-1				3-1					2-1	2-1
UK			1-1	3-1						2-1	2-1

	Road		EMS							
			EMS stations		EMS staff		EMS transportation units		EMS response time	
	Motorway density	% of motorways and national roads in total road length	EMS stations per 1000 km^2	% of EMS stations with at least one physician	EMS medical staff per 10.000 citizens	% of physicians and paramedics	EMS transportation units per 100 km of road length	% of high-equipped transportation units	average response time	% of EMS response meeting the demand
AT	1-1			2-1-1						
BE		1-1				2-1-1	2-1-1			
BG	1-1						2-1-1		2-1-1	
CY	1-1					2-1-1			2-1-1	
CZ	1-1					2-1-1	2-1-1			
DK		1-1				2-1-1		2-1-1		
EE	1-1			2-1-1						2-1-1
FI	1-1			2-1-1						
FR		1-1							2-1-1	
DE		1-1		2-1-1			2-1-1			
EL	1-1		2-1-1							
HU	1-1				2-1-1					
IE	1-1			2-1-1						
IT	1-1			2-1-1			2-1-1			
LV	1-1						2-1-1		2-1-1	
LT	1-1						2-1-1		2-1-1	
LU		2-1		1-1-1			1-1-1			
NL		1-1		2-1-1		2-1-1				
NO	1-1		2-1-1				2-1-1			
PL	1-1				2-1-1					2-1-1
PT	1-1					2-1-1		2-1-1		
RO	1-1					2-1-1				
SK	1-1							2-1-1		
SI		1-1			2-1-1		2-1-1			
ES		1-1				2-1-1				
SE	1-1		2-1-1			2-1-1				
CH		1-1						2-1-1		2-1-1
UK	1-1						2-1-1			

Based on Table 6.4, not only the risk aspects that need urgent policy action for each country can be identified, but some specific road safety enhancing recommendations for all the 28 European countries as a whole can be formulated as well. They are:

- Driving under the influence of alcohol is for most countries the road user behavior risk factor with the highest priority. Although all countries have national policies on drink driving, enforcement remains a critical issue requiring more efforts, especially through increased random breath-testing.
- Speeding is also a major issue for road safety in Europe. Development of effective speed management so as to control the mean speed and the frequency of speed limit violations, especially on urban and rural roads, is of primary importance for the majority of countries.
- With respect to protective systems, more attention should be paid to raise the rear seat belt wearing rates and to improve the monitoring of child restraint system use.
- To improve vehicle active and passive safety performance, countries are first and foremost encouraged to either raise the proportion of new vehicles or reduce the amount of old vehicles in their fleet.
- Relative to vehicle occupant protection and pedestrian protection, improving the safety level on child protection should be an important concern of the vehicle manufacturers. Meanwhile, the installation of seat belt reminders in a vehicle should be greatly advocated in most of the countries.
- In addition, reducing the proportion of powered two-wheelers in the whole vehicle fleet in view of road safety seems to be a more challenging task in Europe compared with the situation of goods vehicles.
- As the most safe type of roads, motorway density is still low in most of the European countries.
- To improve the quality of post-crash medical treatment, the number of EMS transportation units, such as basic life support units, mobile intensive care units and helicopters, as well as the EMS staff, especially the proportion of

physicians and paramedics, have to be focused on in the first place for most countries.

6.6 Conclusion

In this chapter, we investigated the use of data envelopment analysis to develop a composite road safety performance index for cross-country comparison. Starting from the basic DEA-based composite index model, we further explored the incorporation of a layered hierarchy in the DEA framework, and proposed a multiple layer DEA-based composite index model (both the primal and dual form). In general, the model has a similar structure as that of the one layer model except for the additional sets of constraints on layer-specific weights. Thus, the information on the hierarchical structure of the indicators is reflected, and the obtainment of realistic and acceptable indicator weights guaranteed. Moreover, value judgment from decision makers or experts can be easily incorporated by restricting the weight flexibility in a particular category of a layer, which is impossible to be realized in the one layer model. In addition, the extra programming effort is limited and the results can be obtained within a few seconds. However, applying the MLDEA-CI model implies that raw data cannot be used directly. In other words, data must be normalized first in order to remove scale and measurement unit differences.

In the application, the proposed MLDEA-CI model has proven valuable for the road safety context. Above all, the most optimal road safety performance index score was determined for each of the 28 European countries by combining the 32 hierarchical SPIs developed in the previous chapter. Countries were thereby classified into two groups, i.e., best-performing and underperforming. Moreover, the ranking of these countries was deduced by computing their cross-index score, and a clear link with the overall road safety risk from the view of the final outcome level was verified, which in turn justified the use of the proposed multilayer model for composite index construction. Furthermore, rather than postulating the same set of benchmarks (such as the SUN countries) and the same measures for each country, the methodology took the characteristics of each country in the data set into account, and the country-specific benchmarks for those underperforming countries were identified. More importantly, by analyzing the indicator weights allocated in each layer of the hierarchy, useful insight in the areas of underperformance in each country was gained enabling policymakers to prioritize their actions to improve the level of road safety. Learning about best practices applied in benchmarks is therefore a first next step to take based on these results.

However, like any technique, DEA is also characterized by some limitations that need to be kept in mind when interpreting the results. First, the model only measures the performance of one country with respect to the other countries within the sample and a change in the set of countries may lead to other outcomes. Moreover, the results obtained from the DEA model (or the MLDEA-CI model in particular) are also sensitive to indicator specification, hierarchical structure, data quality and chosen weight restrictions. Therefore, as many comparable countries as possible should be considered, appropriate indicators and their structure used, reliable data collected and accepted views from experts adopted to ensure the robustness of the results to an utmost extent. In addition, country comparison over time should be conducted in future research so as to evaluate the results of policy interventions and to monitor the progress in road safety performance.

Chapter 7 Construction of a Composite Index (II): Taking Interval Data into Account

This chapter reconsiders the treatment of missing data in the composite index construction, thereby corresponding to the seventh research question of this dissertation. Instead of using the mean imputed values from multiple imputation, here, missing data are replaced by approximations in the form of intervals in which the true values are believed to lie. An interval MLDEA-based CI model is thereafter used to provide for each country a lower and an upper bound of its performance index score.

7.1 Introduction

In Chapter 6, the hierarchical structure of the safety performance indicators has been considered in the construction of a composite index. This is mainly due to the fact that a large number of SPIs have been developed and further been grouped into categories so as to comprehensively measure the different road safety risk factors in a country. Coupled with the proliferation of SPIs, however, a certain risk in the form of missing values is inevitably present as no data collection system grants perfect data sets. As indicated in Chapter 5, around 23% of the indicator values are missing in this study (see also Figure 5.3). On the other hand, as a 'data-oriented' technique, the applicability of DEA, or the MLDEA-based CI model in particular, relies firstly on the availability of data. In other words, a complete data set with crisp positive values is commonly the prerequisite of the evaluation. To this end, we imputed the missing data by using multiple imputation (MI) in Chapter 5, and all the missing indicator values in the data set were replaced by their *mean* imputed values (see also Table 5.8). In doing so, we implicitly impose an assumption that the average imputations are most likely to correspond with the true values, which however, is not always convincing and can generate a certain degree of bias in the final results. In fact, the application of the MI technique also provides us with an alternative solution for missing values, which is to use imputed data *intervals* [Cherchye et al., 2011]. That is, the missing values are replaced by approximations in the form of intervals estimated from MI in which the true values are believed to lie. As a result, we obtain a complete but imprecise data set that comprises both observed data and imputed data intervals.

In fact, interval data are sometimes also a more logical way to describe the road safety performance of a country with respect to certain aspects, such as the mean speed and the average EMS response time. However, based on such a data set, the MLDEA-based CI model proposed in Chapter 6 cannot be applied directly to compute the composite road safety performance index score for each country. In this chapter, we therefore introduce an interval MLDEA-based CI model building on the contributions of Despotis & Smirlis (2002), Smirlis et al. (2006), and Cherchye et al. (2011). The model is able to provide for each country an upper and a lower bound of its index score corresponding to its most favorable and unfavorable imputation option, respectively. Based on the interval index scores, countries can further be classified, as defined by Cherchye et al. (2011), into 'benchmark countries', 'potential benchmark countries', and 'countries open to improvement'.

The remaining of this chapter is structured as follows. In Section 7.2, we introduce the mechanism of using MI to estimate interval bounds for missing values. In Section 7.3, we formulate the interval MLDEA-based CI model and define upper and lower bound index scores for each country. The application of this model in the construction of a composite road safety performance index is provided in Section 7.4, and the results are given subsequently. The chapter is summarized in Section 7.5.

7.2 Interval Data Generation

As introduced in Section 5.3.2, an important limitation of all the *single imputation* methods for dealing with missing data problems is that they systematically underestimate the variance of the estimates. One solution is to repeat the imputation several times, generating multiple sets of new data whose coefficients vary from set to set. We then capture this variability and add it back into the estimates. This technique is known as *multiple imputation* [Rubin, 1987].

There are a number of ways to perform MI, and the process of MI using a multivariate normal model is relatively straightforward. According to Howell (2008), the first step involves the imputation of a complete set of data from parameter estimates derived from the incomplete data set. Under the multivariate normal model, the imputation of an observation is based on regressing a variable with missing data on the other variables in the data set. Assume, for simplicity, that Y is regressed on only one other variable X. Denote the standard error of the regression as s_{YX}. In standard regression imputation, the imputed value of Y, i.e., $\hat{Y}$, is obtained by: $\hat{Y} = \beta_0 + \beta_1 X$. Moreover, for data augmentation a random error will be added to the prediction by setting $\hat{Y} = \beta_0 + \beta_1 X + \mu s_{YX}$, where μ is a random draw from a standard normal distribution. This way, the necessary level of uncertainty is introduced into the imputed value each time.

Since the parameter estimates such as the regression coefficients and the standard error of regression are all derived from the incomplete data set, and each having its own distribution, the second step is to make a random draw of these estimates from their posterior distribution, i.e., the distribution of the estimates given the data at hand.

Having obtained the initial imputed values and their parameter estimates, the third step is to iterate the process, i.e., imputing values, deriving revised parameter estimates, imputing new values, and so on until convergence is reached. At that point we obtain the final imputed data set.

However, the MI process does not stop yet because only one complete data set is generated. The procedure will therefore start again and generate several more data sets[24]. Because of the randomness inherent in the algorithm, these data sets will differ slightly from one another. Accordingly, when some standard data analysis procedure (e.g., ANOVA) is applied to each data set, the results will be different as well. Once the analyses have been completed for each imputed data set, all that remains is to combine the results to obtain one overall set of estimates following a set of rules provided by Rubin (1987).

[24] Rubin (1987) has shown that in many cases three to five data sets are sufficient.

Specifically, with N imputations, the mean imputed value for Y can be computed as follows:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N}\hat{Y}_i \tag{7-1}$$

With respect to the variance of $\overline{Y}$, two different components can be distinguished. They are the average within-imputation variance $\overline{V}$ and the between-imputation variance B. The first component measures the natural variability in the data, which is analogous to the variance we would produce if we do not need to account for missing data. It can be computed by averaging the variance estimates from each imputed data set ($\hat{V}_i$) as follows:

$$\overline{V} = \frac{1}{N}\sum_{i=1}^{N}\hat{V}_i \tag{7-2}$$

The second component is to capture the extra inferential uncertainty introduced by the existence of missing data. In other words, it measures how the point estimates vary from data set to data set. We can compute this variance by using the following formula:

$$B = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{Y}_i - \overline{Y})^2 \tag{7-3}$$

The total variance (T) associated with $\overline{Y}$ is then a weighted sum of the above two variance components:

$$T = \overline{V} + (1 + \frac{1}{N})B \tag{7-4}$$

According to Rubin (1987), the statistic $(Y - \overline{Y})T^{-1/2}$ is approximately distributed as a Student's *t*-distribution with degrees of freedom:

$$df = (N-1)(1 + \frac{1}{r})^2 \tag{7-5}$$

where r is the between-to-within ratio:

$$r = (1 + \frac{1}{N})\frac{B}{\overline{V}} \tag{7-6}$$

Therefore, confidence intervals can be calculated by taking the overall estimate plus or minus a multiple of the standard error (i.e., the square root of the variance).

Based on the above introduction, at least two types of complete data set can be generated from MI. One is the way applied in Chapter 5, in which all the missing values were replaced by the mean imputed values (after five imputations) calculated from (7-1). In doing so, we implicitly impose the assumption that the average imputations are most likely to correspond with the true values. Its correctness, however, is difficult to judge. Alternatively, by taking one step more, a range of values that constitutes for example a 90% confidence interval around the imputed mean can be derived, and the missing values can then be replaced by an interval within which the true values are believed to lie. That way, the complete data set is composed of both observed data and imputed data intervals.

In this study, we use the multiple imputation procedure in SPSS 17.0 [SPSS Inc., 2007] to generate five complete data sets. Afterwards, the imputed mean and its 90% confidence interval are calculated for each missing value based on formulas (7-1)-(7-6). Taking the mean speed on motorways as an example, the original indicator data and the imputed data derived from the above two methods are illustrated in Table 7.1.

Table 7.1 Original and imputed data for the indicator of mean speed on motorways for the 28 European countries

	Mean speed on motorways		
	Original value	Imputed mean	Imputed interval
AT	0.910	0.910	0.910
BE	1.009	1.009	1.009
BG	0.932	0.932	0.932
CY	1.050	1.050	1.050
CZ	0.835	0.835	0.835
DK	0.936	0.936	0.936
EE	N/A	0.983	[0.946, 1.020]
FI	0.886	0.886	0.886
FR	0.915	0.915	0.915
DE	N/A	0.943	[0.918, 0.969]
EL	N/A	0.916	[0.843, 0.990]
HU	0.858	0.858	0.858

IE	0.903	0.903	0.903
IT	N/A	1.012	[0.910, 1.115]
LV	N/A	0.951	[0.939, 0.963]
LT	0.854	0.854	0.854
LU	0.885	0.885	0.885
NL	0.950	0.950	0.950
NO	1.000	1.000	1.000
PL	N/A	1.059	[1.029, 1.089]
PT	1.008	1.008	1.008
RO	N/A	0.941	[0.817, 1.064]
SK	N/A	0.977	[0.940, 1.015]
SI	0.885	0.885	0.885
ES	0.953	0.953	0.953
SE	0.966	0.966	0.966
CH	0.906	0.906	0.906
UK	0.994	0.994	0.994

As can be seen, eight countries have no available data for this indicator. They are Estonia, Germany, Greece, Italy, Latvia, Poland, Romania, and Slovakia. Among the observed data, Cyprus has the highest indicator value (1.05), or the worst performance, while Czech Republic has the best (0.835). Now, by taking all the indicators belonging to the road user behavior into account for missing data imputation, Poland obtains the highest imputed mean (1.059), indicating that it has even worse performance than Cyprus with regard to this risk aspect. However, no country has an imputed indicator value lower than that of Czech Republic, indicating that none of them can overtake Czech Republic in this respect. Nevertheless, if we use interval data for replacement, such a crisp judgment is somewhat nuanced. For instance, Poland is no longer the worst country in all the cases. Although the performance of Poland is still poor compared to most of the countries even when its best possible value (1.029) is considered, countries like Italy and Romania could perform even worse if they catch the upper bound of their intervals. On the other hand, if the lower bound of the interval is reached, Romania can also become the best-performer in this aspect. Although ambiguity, using the interval data is more in accordance with our intuition on missing data, and more acceptable to countries and policy makers. Moreover, the relative big interval ranges derived for countries such as

Romania and Italy can be further interpreted by the fact that they owns the greatest number of missing values in the data set (17 and 16, respectively), which renders their five imputed values to vary considerably from one data set to another. In other words, uncertainty due to the imputation is relatively high for these two countries. This information, however, cannot be reflected by using the mean imputed values.

7.3 Interval MLDEA-based CI model

Due to the mixture of observed data and interval data, the MLDEA-based CI model (re-presented in (7-7)) is no longer linear as, apart from the original variables $\hat{u}_1, \ldots, \hat{u}_{f_1}, \ldots \hat{u}_s$ (i.e., the indicator weights), the indicators themselves, i.e., $y_{f_1 j}$, are also variables whose exact values are not known, but lying within bounded intervals $[y^{\mathrm{L}}_{f_1 j}, y^{\mathrm{U}}_{f_1 j}]$.

$$
\begin{aligned}
CI_c &= \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c} \\
s.t. \quad & \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j} \le 1, \quad j = 1, \cdots, n \\
& \sum_{f_1 \in A^{(k)}_{f_k}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A^{(k+1)}_{f_{k+1}}} \hat{u}_{f_1} = \underset{f_k \in A^{(k+1)}_{f_{k+1}}}{p^{(k)}_{f_k}} \in \Theta, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \\
& \hat{u}_{f_1} \ge 0, \quad f_1 = 1, \cdots, s
\end{aligned}
\tag{7-7}
$$

Despotis & Smirlis (2002) suggested the following transformation to convert the non-linear model (7-7) into a linear one:

$$
y_{f_1 j} = y^{\mathrm{L}}_{f_1 j} + t_{f_1 j}(y^{\mathrm{U}}_{f_1 j} - y^{\mathrm{L}}_{f_1 j}), \quad f_1 = 1, \cdots, s, \quad j = 1, \cdots, n, \quad 0 \le t_{f_1 j} \le 1 \tag{7-8}
$$

By using this expression, the term $\hat{u}_{f_1} y_{f_1 j}$ is replaced by $\hat{u}_{f_1} y^{\mathrm{L}}_{f_1 j} + \hat{u}_{f_1} t_{f_1 j}(y^{\mathrm{U}}_{f_1 j} - y^{\mathrm{L}}_{f_1 j})$. We then introduce a new variable $q_{f_1 j} = \hat{u}_{f_1} t_{f_1 j}$ which meets the condition $0 \le q_{f_1 j} \le \hat{u}_{f_1}, \ \forall f_1, j$.

Applying the above transformations to model (7-7), we obtain the following linear programming problem:

$$
\begin{aligned}
CI_c &= \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c}^{\mathrm{L}} + q_{f_1 c}(y_{f_1 c}^{\mathrm{U}} - y_{f_1 c}^{\mathrm{L}}) \\
s.t. \quad & \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j}^{\mathrm{L}} + q_{f_1 j}(y_{f_1 j}^{\mathrm{U}} - y_{f_1 j}^{\mathrm{L}}) \le 1, \quad j = 1, \cdots, n \\
& \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{p_{f_k}^{(k)}} \in \Theta, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \\
& q_{f_1 j} - \hat{u}_{f_1} \le 0, \quad f_1 = 1, \cdots, s, \quad j = 1, \cdots, n \\
& \hat{u}_{f_1} \ge 0, \quad f_1 = 1, \cdots, s
\end{aligned}
\tag{7-9}
$$

In fact, model (7-7) is a special case of model (7-9) in which all the lower and upper bounds coincide for all the indicators. In this case, exact rather than interval data are actually used for the calculation. The variable $q_{f_1 j}$ is then eliminated and model (7-9) is reduced to model (7-7). When interval data exist, i.e., the lower and upper bounds are not identical for all the indicators, the optimal index score of country *c* is obtained by adjusting not only the weights but also the levels of indicators within their ranges that are in favor of it. In other words, the index score attained by country *c* in model (7-9) is not worse than any other index score that the country might attain, by adjusting the indicator values in a different way within the limits of the bounded intervals. Smirlis et al. (2006) further proved that such an optimal index score can be obtained from the following model with exact data:

$$
\begin{aligned}
CI_c &= \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c}^{\mathrm{U}} \\
s.t. \quad & \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c}^{\mathrm{U}} \le 1, \\
& \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j}^{\mathrm{L}} \le 1, \quad j = 1, \cdots, n, \quad j \ne c \\
& \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{p_{f_k}^{(k)}} \in \Theta, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \\
& \hat{u}_{f_1} \ge 0, \quad f_1 = 1, \cdots, s
\end{aligned}
\tag{7-10}
$$

In model (7-10), the country under evaluation is set in its best possible position (i.e., the indicator values are all adjusted to their upper bound) while all the other countries in the data set are set in their least favorable position (i.e., the indicator values are contrarily adjusted to their lower bound). Cherchye et al. (2011) defined it as a ‘strong country in a weak environment’ scenario. We thus obtain the upper bound of the possible index score that country *c* might attain in an interval data setting (referred to as CI_c^{upper}).

Likewise, a lower bound of the index score for country *c* can be obtained from model (7-11) below.

$$
\begin{aligned}
CI_c &= \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c}^{\mathrm{L}} \\
s.t. \quad & \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 c}^{\mathrm{L}} \le 1, \\
& \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j}^{\mathrm{U}} \le 1, \quad j = 1, \cdots, n, \quad j \ne c \\
& \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Bigg/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{p_{f_k}^{(k)}} \in \Theta, \quad f_k = 1, \cdots, s^{(k)}, \quad k = 1, \cdots, K-1 \\
& \hat{u}_{f_1} \ge 0, \quad f_1 = 1, \cdots, s
\end{aligned}
\tag{7-11}
$$

Model (7-11) is also a MLDEA-based CI model with exact data. In contrast to model (7-10), the levels of indicators are now adjusted unfavorably for the evaluated country *c* (i.e., the indicator values are set to their lower bound) and in favor of the other countries in the data set (i.e., the indicator values are set to their upper bound). Cherchye et al. (2011) defined this as a 'weak country in a strong environment' scenario, and the model results in a lower bound of the possible index score for country *c* (i.e., CI_c^{lower}).

Models (7-10) and (7-11) constitute the interval MLDEA-based CI model, which therefore provides for each country a bounded interval of its index score $[CI_c^{lower}, CI_c^{upper}]$, within which the exact one is believed to lie. Moreover, the length of the obtained interval reflects the overall uncertainty due to the underlying imperfect nature of the indicator data.

7.4 Application and Results

Using the interval indicator data generated in Section 7.2, we now apply the interval MLDEA-based CI model to combine all the SPIs into a composite road safety performance index for the 28 European countries. Given the same data normalization procedure and weight restrictions as in Chapter 6, the results based on models (7-10) and (7-11) are presented in Table 7.2, together with the ones obtained in Chapter 6 (referred to as CI_c^*).

Table 7.2 Composite index scores derived from exact and interval data

	CI_c^{lower}	CI_c^{upper}	CI_c^*
AT	0.883	1.000	0.965
BE	0.847	1.000	0.958
BG	0.730	1.000	0.869
CY	0.680	0.928	0.834
CZ	0.845	0.960	0.912
DK	0.760	1.000	0.884
EE	0.857	1.000	0.928
FI	0.816	0.966	0.906
FR	0.866	1.000	0.944
DE	0.909	1.000	1.000
EL	0.601	0.885	0.749
HU	0.657	0.794	0.751
IE	0.773	0.987	0.905
IT	0.809	1.000	0.967
LV	0.647	0.863	0.746
LT	0.687	0.844	0.805
LU	1.000	1.000	1.000
NL	0.978	1.000	1.000
NO	0.874	0.981	0.941
PL	0.805	0.858	0.845
PT	0.766	1.000	0.903
RO	0.547	0.988	0.766
SK	0.692	1.000	0.877
SI	0.761	1.000	0.913
ES	0.857	1.000	0.961
SE	1.000	1.000	1.000
CH	0.902	1.000	1.000
UK	0.906	0.986	0.971

By taking the missing data uncertainty into account, an imprecise index score in the form of a bounded interval $[CI_c^{lower}, CI_c^{upper}]$ is obtained for each country based on the interval MLDEA-based CI model, within which the precise index score CI_c^* calculated in Chapter 6 is always situated (see also Figure 7.1). In other words, CI_c^* only shows one possibility of a country's overall road safety performance in the context of missing data, and theoretically, it can be switched into any value within the interval range. Consequently, the interval index score provides us with a more credible representation of a country's overall road safety performance as it highlights rather than conceals the underlying imperfect nature of the indicator data. Under this circumstance, however, countries can no longer be fully ranked as was the case based on the precise index scores. The most we can judge from Table 7.2 is that the overall road safety performance of Hungary ([0.657, 0.794]) is worse than that of countries like Poland ([0.805, 0.858]), which performs in turn worse than countries like Norway ([0.874, 0.981]), and so on. Moreover, Luxembourg and Sweden are the only two countries with their index score always equal to one, no matter which scenario is taken into account. In other words, the influence of the existence of missing data in the data set on the final index score of these two countries can be ignored, given the pre-specified confidence level. Therefore, other countries cannot dispute that they are the best-performers among these countries, and they can be unambiguously designated as 'benchmark countries'.

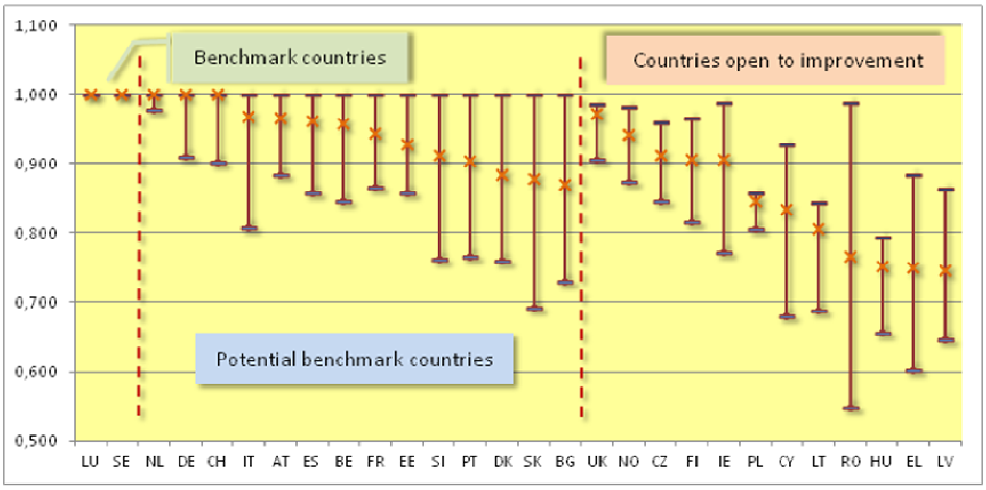


Figure 7.1 Country classification based on the interval MLDEA-based CI model

With regard to the remaining countries, based on their best possible index score, i.e., CI_c^{upper}, countries can be further classified, as defined by Cherchye et al.

(2011), into 'potential benchmark countries' and 'countries open to improvement'. The latter are those countries who can only obtain an optimal index score less than one even when the extreme 'strong country in weak environment' scenario is taken into account. Therefore, countries belonging to this classification are definitely underperforming, and they cannot complain about data problems. Some ambiguity, however, is associated with those 'potential benchmark countries', for which the actual classification as a best performer or not is contingent on the actual imputation of specific indicator values. Nevertheless, as indicated before, the length of the obtained interval for each country reflects the overall uncertainty about its index score due to missing data. The Netherlands, for instance, obtains relatively high index scores in both scenarios with a small difference between them (1-0.978=0.022), indicating that the influence of missing data in the data set on its index score is limited, and it is more tending to be assigned as a best performer (with the real index score of one), and this is verified by its CI_c^*. On the contrary, countries like Slovakia and Bulgaria obtain large interval ranges of their index score, implying that much ambiguity remains about their index score, and they are more likely to be the 'countries open to improvement'. The overall classification of the 28 European countries is illustrated in Figure 7.1. Countries in each classification are ranked by CI_c^* (pointed out with the crosses).

7.5 Conclusion

In this chapter, we investigated the influence of the existence of missing data in the data set on the final index score of 28 European countries. In doing so, missing data were firstly replaced by approximations in the form of intervals deduced from multiple imputation in which the true values are believed to lie. Thus, the complete data set consisted of both observed data and imputed data intervals. The MLDEA-based CI model was accordingly extended in order to take interval data into account. Its application resulted in an upper and a lower bound of the index score for each country corresponding to the most favorable and unfavorable option, respectively. The index interval instead of the precise index score for each country highlights the underlying imperfect nature of the indicator data, and provides us with a more credible representation of a country's overall road safety performance. Based on the interval index scores, countries were further classified into 'benchmark countries', 'potential benchmark countries', and 'countries open to improvement'. With respect to those benchmark countries, such as Luxembourg and Sweden in this study, the influence of missing data in the data set on their index score can be ignored, for the pre-specified confidence level. While for the third class of countries requiring further improvement in their road safety performance, missing indicator values play a certain role in determining their optimal index score, but have no implication in classifying them as underperforming countries. Finally, concerning those 'potential benchmark countries', the actual classification as a best performer or not is to a great extent determined by the actual imputation of specific indicator values. However, a large interval range around the index score is generally derived for those countries whose missing data uncertainty is relatively high, and such countries are much more likely to be the best performer only by coincidence.

Chapter 8 Construction of a Composite Index (III): Modeling Qualitative Data[25]

This chapter focuses on the last research question of this dissertation, which is to model qualitative data in the context of composite index construction. Two strategies, i.e., the imprecise DEA-based CI model and the fuzzy DEA-based CI model are thereby investigated in this chapter. The models are demonstrated by taking the qualitative alcohol indicator developed in Chapter 5 into account. The fuzzy MLDEA-based CI model is further applied to create a composite road safety performance index.

8.1 Introduction

In the previous chapters, the construction of a composite index (CI), using either a multilayer model or an interval model, has generally been assumed to be based upon a set of quantitative indicators. Under many conditions, however, quantitative data are inadequate or inappropriate to model real world situations due to the complexity and uncertainty of the reality. Therefore, it is essential to also take the presence of qualitative indicators into account when making a decision on the performance of a country. Very often it is the case that an indicator can, at most, be specified with either ordinal measures, from best to worst, or with the help of experts' subjective judgments, such as 'high', 'medium' and 'low'. In the development of alcohol indicators in Chapter 5, apart from the safety performance indicator, i.e., *the percentage of drivers above the legal BAC limit in roadside checks*, and the indicator representing the consequence of drink driving from the view of the final outcome level, i.e., *the percentage of road fatalities attributed to alcohol*, an additional indicator related to policy output, i.e., *the effectiveness of the overall enforcement against drinking and driving*, was also suggested to supplement the alcohol performance

[25] Related research has been published in: Shen, Y., Ruan, D., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K., (2011). Modeling qualitative data in data envelopment analysis for composite indicators, *International Journal of Systems Assurance Engineering and Management*, Vol. 2, No. 1, pp. 21-30.

of a country. Such a policy performance indicator, however, is qualitative in nature, and only takes the form of ordered classes rated on a 0-10 scale rather than numerical values for the purpose of describing, comparing and evaluating this risk factor among various countries. To include such kind of indicators in the construction of a composite road safety performance index, the DEA-related models developed so far in the previous chapters are not capable of deriving a satisfactory solution.

Generally, two strategies have appeared in the literature for the treatment of qualitative data within the DEA framework. One is to reflect the rank position of each DMU with respect to each ordinal indicator by setting corresponding constraints, which is collectively referred to as imprecise DEA (IDEA). Cook et al. (1993, 1996) first presented this kind of DEA structure and applied it to the problem of prioritizing a set of research and development projects. Cooper et al. (1999, 2002) proposed an alternative approach also using this strategy, with an illustrative application to a Korean mobile telecommunication company. These two approaches to the treatment of ordinal data in DEA were further discussed and compared in Cook & Zhu (2006, 2007).

The other strategy is to deal with the natural uncertainty inherent to some production processes by means of fuzzy mathematical programming, which leads up to the so-called fuzzy DEA (FDEA). We can find several fuzzy approaches for evaluating DMUs in the DEA literature. The tolerance approach was one of the first FDEA models that was developed by Sengupta (1992) and further improved by Kahraman & Talgo (1998). The main idea of this approach was to incorporate uncertainty into the DEA models by defining tolerance levels on constraint violations. Lertworasirikul (2001) developed a defuzzification approach, in which the fuzzy inputs and outputs are first defuzzified into crisp values, and then the conventional DEA model was applied. Meada et al. (1998) introduced an α-level based approach, which was further developed by Saati et al. (2002). In this method, a FDEA model was solved by parametric programming using α-cuts. At a given level of α-cut, it produced the interval efficiency for the DMU under assessment. Kao & Liu (2000) proposed a method to find the membership functions of the fuzzy efficiency scores when some observations were fuzzy numbers. The idea was based on the α-cuts and Zadeh

(1965)'s extension principle. Lertworasirikul et al. (2003) proposed a possibility approach in which constraints are treated as fuzzy events. The approach transforms FDEA models into possibility DEA models by using possibility measures of fuzzy constraints. Entani et al. (2002) and Wang et al. (2005) used the concept of interval efficiency assessment in DEA. Saati & Memariani (2005) employed the case of reducing weight flexibility in FDEA. Hougaard (2005) introduced a simple approximation of productivity scores of a fuzzy production plan, which allowed the decision makers to use scores of technical efficiency in combination with other sources of information such as expert opinions. Guo & Tanaka (2001) proposed a fuzzy ranking approach, in which both fuzzy inequalities and fuzzy equalities were defined by ranking methods and the resulting model was based on a bi-level linear programming and provided fuzzy efficiency for an evaluated DMU at a specified α-cut. The approach was applied to a restaurant location problem in Guo (2009). Based on a similar manner, León et al. (2003) extended the models from the primal multiplier formulation to the dual envelopment one, and Tlig & Rebai (2009) further transformed the models into crisp linear programming problems, which produced crisp efficiency, rather than fuzzy efficiency.

In the remaining sections of this chapter, we investigate both strategies for modeling qualitative data in the construction of CIs. More specifically, based on the imprecise DEA introduced in Cook & Zhu (2007) and the fuzzy ranking approach proposed in Guo (2009), we develop two new models for composite index construction in Section 8.2 and Section 8.3, respectively. We illustrate these two models with the application of constructing a composite alcohol performance index in Section 8.4. In Section 8.5, a fuzzy multiple layer DEA-based CI model is further developed to combine all the 33 hierarchical indicators (with both quantitative and qualitative data) into a composite road safety performance index. The chapter ends with conclusions in Section 8.6.

8.2 Imprecise DEA for CIs

In the basic DEA-based CI model (re-presented in (8-1)), obtainment of measurable and quantitative indicators is commonly the prerequisite of the evaluation.

$$
\begin{aligned}
CI_c &= \max \sum_{r=1}^{s} u_r y_{rc} \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{rj} \le 1, \quad j = 1,\cdots,n \\
& u_r \ge \varepsilon, \quad r = 1,\cdots,s
\end{aligned}
\tag{8-1}
$$

In situations where some indicators might better be represented as rank positions in an ordinal, rather than numerical sense, this model cannot be used directly, because ordinal data cannot be simply treated as numerical ones for which a score of 2 is twice as large as a score of 1. The most that can be judged is that the former one is preferred to or more important than the latter in a maximization context. Therefore, in combining all the indicators into one index score, the numerical and ordinal ones need to be treated differently. In this section, we introduce imprecise DEA for modeling ordinal data in the construction of CIs.

Consider a set of n DMUs that is to be evaluated in terms of s_1 numerical indicators (y^n) and s_2 ordinal indicators (y^o). The modified DEA-based CI model can then be formulated as follows:

$$
\begin{aligned}
CI_c &= \max \sum_{r=1}^{s_1} u_r y_{rc}^n + \sum_{i=1}^{s_2} v_i y_{ic}^o \\
s.t. \quad & \sum_{r=1}^{s} u_r y_{rj}^n + \sum_{i=1}^{s_2} v_i y_{ij}^o \le 1, \quad j = 1,\cdots,n \\
& u_r, v_i \ge \varepsilon, \quad r = 1,\cdots,s_1, \ i = 1,\cdots,s_2
\end{aligned}
\tag{8-2}
$$

Since precise values for indicators y_{ij}^o ($i = 1,\cdots,s_2$) are not available, but the ordinal data with a K-point scale, we first convert the ranking value of each DMU into its position value. Taking K=4 as an example, the conversion of the four ordinal numbers is illustrated in Table 8.1.

Table 8.1 Conversion of 4-point scale ordinal data

Ordinal data (y_{ij}^o)		Position values (p_{ki})			
1		0	0	0	1
2	⇨	0	0	1	0
3		0	1	0	0
4		1	0	0	0

Each position vector p_{ki} ($k = 1, \cdots, K$) is then multiplied by a vector $[y_{i1}, y_{i2}, \cdots y_{iK}]^T$ to obtain the virtual value of each ordinal indicator, i.e., $y_i^v = p_{ki}[y_{i1}, y_{i2}, \cdots y_{iK}]^T$, which should satisfy the minimal requirement that $y_{i1} > y_{i2} > \cdots > y_{iK} \geq \varepsilon > 0$. Thus, the aggregation of all the s_2 ordinal indicators for a particular country c can be expressed as: $\sum_{i=1}^{s_2} v_i y_{ic}^v$, which can be further linearized as $\sum_{i=1}^{s_2} \sum_{k=1}^{K} \mu_{ikc}$. Moreover, to realize the above ordinal restrictions on y_{ik} ($k = 1, \cdots, K$), we set $y_{ik} - y_{ik+1} \geq \varepsilon$ [26] $(k = 1, \cdots K - 1)$, and $y_{iK} \geq \varepsilon$. In this way, $\mu_{ik} - \mu_{ik+1}$ which equals to $v_i y_{ik} - v_i y_{ik+1}$ should at least satisfy the minimum constraint that $\mu_{ik} - \mu_{ik+1} \geq \varepsilon^2$, and $\mu_{rK} \geq \varepsilon^2$. We thus obtain the IDEA-based CI model as follows:

$$
\begin{aligned}
& CI_c = \max \sum_{r=1}^{s_1} u_r y_{ic} + \sum_{i=1}^{s_2} \sum_{k=1}^{K} \mu_{ikc} \\
& s.t. \quad \sum_{r=1}^{s_1} u_r y_{rj} + \sum_{i=1}^{s_2} \sum_{k=1}^{K} \mu_{ikj} \leq 1, \quad j = 1, \cdots, n \\
& \qquad u_r \geq \varepsilon, \quad r = 1, \cdots, s_1 \\
& \qquad \mu_{ik} - \mu_{ik+1} \geq \varepsilon^2, \quad k = 1, \cdots, K-1, \quad i = 1, \cdots, s_2 \\
& \qquad \mu_{iK} \geq \varepsilon^2, \quad i = 1, \cdots, s_2
\end{aligned}
\tag{8-3}
$$

In model (8-3), both numerical and ordinal indicators are taken into account, and due to the incorporation of ordinal indicators, two additional inequality constraints are added as shown above.

8.3 Fuzzy DEA for CIs

To treat imprecision and vagueness in a DEA model, apart from IDEA, fuzzy set theory can also be integrated as an alternative solution. By interpreting the

[26] ε is a small number used to reflect the minimum allowable gap between the values associated with y_{ik} and y_{ik+1}, which can have a certain impact on the final index scores. In real situations, different ε values can be used for different ordinal indicators, or other discrimination intensity functions can be employed. See also Cook & Kress (1990).

qualitative (or ordinal) data as fuzzy numerical values which can be represented by means of fuzzy numbers or fuzzy intervals, the basic DEA-based CI model (8-1) can be naturally extended to the following fuzzy one:

$$\begin{aligned} CI_c &= \max \sum_{r=1}^{s} u_r \tilde{y}_{rc} \\ s.t. \quad & \sum_{r=1}^{s} u_r \tilde{y}_{rj} \lesssim 1, \quad j = 1, \cdots, n \\ & u_r \geq \varepsilon, \quad r = 1, \cdots, s \end{aligned} \tag{8-4}$$

where $\tilde{y}_{rj}$ denotes the *r*th fuzzy indicator value of the *j*th country.

The resulted FDEA-based CI model (8-4) takes the form of a fuzzy linear programming problem with fuzzy coefficients in the objective function and also the constraints. Therefore, some fuzzy operations including 'maximizing a fuzzy variable' and 'fuzzy inequality' are required. In what follows, we simply recall how to perform the basic operations of arithmetics and the comparison of fuzzy intervals for ranking purposes. To be more precise, we deal with *LR*-fuzzy numbers whose definition is as follows.

Definition 1 [León et al., 2003]. A fuzzy number $\tilde{M}$ is an *LR*-fuzzy number, $\tilde{M} = (m^L, m^R, a^L, a^R)_{L,R}$, if its membership function has the following form:

$$\mu_{\tilde{M}}(r) = \begin{cases} L\left(\dfrac{m^L - r}{a^L}\right), & r \leq m^L \\ 1, & m^L \leq r \leq m^R \\ R\left(\dfrac{r - m^R}{a^R}\right), & r \geq m^R \end{cases} \tag{8-5}$$

where the subset $[m^L, m^R]$ consists of the real numbers with the highest chance of realization, a^L is the left spread, a^R is the right spread, and L and R are reference functions defining the left and the right shapes of the fuzzy number, respectively, which should satisfy the following conditions:

$L, R : [0, +\infty) \rightarrow [0,1]$,

$L(x) = L(-x),\ R(x) = R(-x)$,

$L(0) = 1,\ R(0) = 1$, and

$L(x)$ and $R(x)$ are strictly decreasing and upper semi-continuous on $supp(\tilde{M}) = \{r : \mu_{\tilde{M}}(r) > 0\}$.

In addition, an *LR* fuzzy number becomes an *LL* fuzzy number when $L(x) = R(x)$, an *LL* fuzzy number with $L(x) = \max(0, 1 - |x|)$ is known as a triangular fuzzy number, and a symmetrical *LL* fuzzy number is for the case of $\alpha^L = \alpha^R$.

Let us now recall the definition of the maximum of two fuzzy numbers.

Definition 2 [Inuiguchi et al., 1990]. Let $\tilde{M}$ and $\tilde{N}$ be two fuzzy numbers and h a real number, $h \in [0,1]$. Then $\tilde{M} \gtrsim^h \tilde{N}$ if and only if, $\forall k \in [h,1]$, the following two statements hold:

$$\begin{aligned} &inf\{s : \mu_{\tilde{M}}(s) \geq k\} \geq inf\{t : \mu_{\tilde{N}}(t) \geq k\} \\ &sup\{s : \mu_{\tilde{M}}(s) \geq k\} \geq sup\{t : \mu_{\tilde{N}}(t) \geq k\} \end{aligned} \tag{8-6}$$

where *inf* stands for infimum (lower bound or minimum), and *sup* stands for supremum (upper bound or maximum).

Hence, for *LR*-fuzzy numbers with bounded support, and using this ranking method, at a given possibility level *h*, expression (8-6) becomes

$$\begin{aligned} &m^L - L^*(k)\alpha^L \geq n^L - L'^*(k)\beta^L \quad \forall k \in [h,1] \\ &m^R + R^*(k)\alpha^R \geq n^R + R'^*(k)\beta^R \quad \forall k \in [h,1] \end{aligned} \tag{8-7}$$

Therefore, using *LR* fuzzy numbers in the FDEA-based CI model (8-4), i.e., $\tilde{y}_{rj} = (y_{lrj}, y_{urj}, a_{rj}, b_{rj})$, the constraint $\sum_{r=1}^{s} u_r \tilde{y}_{rj} \lesssim 1$ can then be considered as inequalities between an *LR* fuzzy number and a real number, and the use of an ordering relation in (8-7) allows us to convert this fuzzy constraint into a crisp inequality as: $\sum_{r=1}^{s} u_r \left(y_{urj} + b_{rj} R^*(h)\right) \leq 1$ [27].

[27] $\sum_{r=1}^{s} u_r \left(y_{lrj} - a_{rj} L^*(h)\right) \leq 1$ is always satisfied when $\sum_{r=1}^{s} u_r \left(y_{urj} + b_{rj} R^*(h)\right) \leq 1$.

Concerning 'maximizing a fuzzy variable', i.e., $\max \sum_{r=1}^{s} u_r \tilde{y}_{rc}$, still using the ordering relation in (8-7), this objective function can then be decomposed into two crisp relations as: $\max \sum_{r=1}^{s} u_r \left(y_{lrc} - a_{rc} L_{rc}^{*}(h) \right)$ and $\max \sum_{r=1}^{s} u_r \left(y_{urc} + b_{rc} R_{rc}^{*}(h) \right)$, $h \in [0,1]$, which should be maximized simultaneously. To this end, a weighted function $\lambda_1 \sum_{r=1}^{s} u_r \left(y_{lrc} - a_{rc} L_{rc}^{*}(h) \right) + \lambda_2 \sum_{r=1}^{s} u_r \left(y_{urc} + b_{rc} R_{rc}^{*}(h) \right)$ with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, and $\lambda_1 + \lambda_2 = 1$ is used to obtain the compromise solution. Three situations are usually considered, which are optimistic if $\lambda_2 = 1$, pessimistic if $\lambda_1 = 1$, and indifferent if $\lambda_1 = \lambda_2$.

Thus, the FDEA-based CI model (8-4) can now be transformed in the following crisp linear programming problem:

$$
\begin{aligned}
CI_c &= \max \lambda_1 \sum_{r=1}^{s} u_r \left(y_{lrc} - a_{rc} L_{rc}^{*}(h) \right) + \lambda_2 \sum_{r=1}^{s} u_r \left(y_{urc} + b_{rc} R_{rc}^{*}(h) \right) \\
s.t. \quad & \sum_{r=1}^{s} u_r \left(y_{urj} + b_{rj} R_{rj}^{*}(h) \right) \leq 1, \quad j = 1, \cdots, n \\
& u_r \geq \varepsilon, \quad r = 1, \cdots, s
\end{aligned} \tag{8-8}
$$

Definition 3. Country *c* is called fuzzy best performing if and only if it obtains a fuzzy index score of one at least at one possibility level *h*. Otherwise, it is fuzzy underperforming.

Definition 4. Country *c* is called fuzzy non-dominated best performing if and only if it obtains a fuzzy index score of one at all possibility levels *h*.

In particular, if indicators $\tilde{y}_{rj}$ are assumed to be symmetrical triangular fuzzy numbers, which are often used to represent the uncertainty of information for simplification, they can then be denoted by the pairs consisting of the corresponding centers and spreads, $\tilde{y}_{rj} = (y_{rj}, a_{rj})$, $r = 1, \cdots, s$, $j = 1, \cdots, n$, and the model (8-8) can be substantially simplified as follows:

$$
\begin{aligned}
CI_c &= \max \lambda_1 \sum_{r=1}^{s} u_r \left(y_{rc} - (1-h)a_{rc}\right) + \lambda_2 \sum_{r=1}^{s} u_r \left(y_{rc} + (1-h)a_{rc}\right) \\
s.t. \quad & \sum_{r=1}^{s} u_r \left(y_{urj} + (1-h)a_{rj}\right) \le 1, \quad j = 1, \cdots, n \qquad (8\text{-}9) \\
& u_r \ge \varepsilon, \quad r = 1, \cdots, s
\end{aligned}
$$

Note that for triangular fuzzy numbers, $L_{rj}^{*}(h) = R_{rj}^{*}(h) = 1 - h,\ 0 \le h \le 1,\ r = 1, \cdots s$.

The fuzzy index score of country c can then be defined as $\{ \sum_{r=1}^{s} u_r^{*} \left(y_{rc} - (1-h)a_{rc}\right)$, $\sum_{r=1}^{s} u_r^{*} y_{rc}$, $\sum_{r=1}^{s} u_r^{*} \left(y_{rc} + (1-h)a_{rc}\right)$ }, which represents the pessimistic, indifferent, and optimistic situation, respectively.

8.4 Application to a Composite Alcohol Performance Index

To illustrate the use of these two models developed in the above two sections, we apply them for constructing a composite alcohol performance index based on both quantitative and qualitative indicators. More specifically, in the development of alcohol indicators in Chapter 5, apart from one safety performance indicator, i.e., *the percentage of drivers above the legal BAC limit in roadside checks*, and one indicator representing the consequence of drink driving from the view of the final outcome level, i.e., *the percentage of road fatalities attributed to alcohol*, a third indicator related to policy output, i.e., *the effectiveness of overall enforcement against drinking and driving*, was also suggested to supplement the alcohol performance of a country. Such a policy performance indicator, derived from the Global Status Report on Road Safety prepared by the World Health Organization (2009), in which the respondents were asked to reach a consensus on their assessment of the enforcement in the country, is qualitative in nature, and can only take the form of ordered classes rated on a 0-10 scale (with 0 represents the worst drink driving enforcement while 10 the best) rather than numerical values for the purpose of description, comparison and evaluation of this risk factor for various countries. Data on this qualitative indicator for the 28 European countries[28] are presented in Table 8.2, together with the normalized data on the first two quantitative indicators.

Table 8.2 Normalized numerical data and ordinal data on three alcohol indicators for 28 European countries

	Alcohol		
	% of drivers above legal alcohol limit in roadside checks	% of alcohol-related fatalities	Effectiveness of overall enforcement on drinking and driving
AT	0.116	0.463	9
BE	0.068	0.654	3
BG	0.123	0.855	7
CY	0.137	0.182	4
CZ	0.145	0.675	9
DK	0.301	0.143	8
EE	0.860	0.080	8
FI	0.593	0.136	8
FR	0.263	0.123	4
DE	0.093	0.306	4
EL	0.273	0.432	7
HU	0.279	0.283	5
IE	0.237	0.119	5
IT	0.098	0.992	7
LV	0.218	0.175	7
LT	0.555	0.321	6
LU	0.102	0.248	5
NL	0.081	1.000	9
NO	0.142	0.159	4
PL	0.091	0.438	7
PT	0.137	0.610	8
RO	0.070	0.423	8
SK	0.067	0.607	9
SI	0.122	0.078	6
ES	0.398	0.402	7
SE	1.000	0.357	6
CH	0.141	0.230	6
UK	0.051	0.228	5

[28] Data for Ireland, the Netherlands, and United Kingdom are imputed values derived from the multiple imputation procedure in SPSS 17.0.

To combine these three alcohol indicators into one index score, the IDEA-based CI model (8-3) can be applied directly, with the ε value chosen as 0.0001. Whereas for the FDEA-based CI model (8-9), symmetrical triangular fuzzy numbers are used for the ordinal data in this study, which are defined as in Table 8.3.

Table 8.3 Representation of symmetrical triangular fuzzy numbers for the ordinal indicator values

Ordinal data ($\tilde{y}_{rj}$)	Symmetrical triangular fuzzy numbers (y_{rj}, a_{rj})	Ordinal data ($\tilde{y}_{rj}$)	Symmetrical triangular fuzzy numbers (y_{rj}, a_{rj})
0	$\left(0, \frac{1}{10}\right)$	1	$\left(\frac{1}{10}, \frac{1}{10}\right)$
2	$\left(\frac{2}{10}, \frac{1}{10}\right)$	3	$\left(\frac{3}{10}, \frac{1}{10}\right)$
4	$\left(\frac{4}{10}, \frac{1}{10}\right)$	5	$\left(\frac{5}{10}, \frac{1}{10}\right)$
6	$\left(\frac{6}{10}, \frac{1}{10}\right)$	7	$\left(\frac{7}{10}, \frac{1}{10}\right)$
8	$\left(\frac{8}{10}, \frac{1}{10}\right)$	9	$\left(\frac{9}{10}, \frac{1}{10}\right)$
10	$\left(1, \frac{1}{10}\right)$		

In addition, to guarantee that all the three indicators will be used to some extent by the models, the share of each of these three indicators in the final index score is restricted to lie within the interval [0.1, 0.5], yet is rather broad to allow a high level of flexibility.

The composite alcohol performance index score of the 28 European countries can now be computed based on the two models, respectively. The results are shown in Table 8.4.

Table 8.4 Composite alcohol performance index scores of 28 European countries based on the IDEA-based CI model and the FDEA-based CI model

IDEA-CI		FDEA-CI									
			h=0			*h*=0.5			*h*=1		
SE	1.000	SE	{.872,	.947,	1.000}	{.940,	.973,	1.000}	{1.000,	1.000,	1.000}
CZ	.880	CZ	{.768,	.792,	.812}	{.795,	.806,	.816}	{.820,	.820,	.820}
ES	.847	ES	{.684,	.733,	.775}	{.729,	.752,	.774}	{.773,	.773,	.773}
FI	.833	LT	{.670,	.727,	.778}	{.721,	.750,	.776}	{.774,	.774,	.774}
PT	.826	PT	{.694,	.727,	.749}	{.726,	.740,	.752}	{.755,	.755,	.755}
LT	.803	FI	{.686,	.720,	.750}	{.719,	.735,	.751}	{.752,	.752,	.752}
EL	.780	BG	{.672,	.703,	.729}	{.703,	.717,	.730}	{.732,	.732,	.732}
BG	.776	EL	{.634,	.679,	.717}	{.674,	.696,	.715}	{.713,	.713,	.713}
AT	.711	AT	{.624,	.642,	.658}	{.645,	.654,	.662}	{.666,	.666,	.666}
IT	.679	IT	{.598,	.623,	.643}	{.623,	.634,	.644}	{.646,	.646,	.646}
NL	.678	NL	{.566,	.579,	.590}	{.581,	.587,	.592}	{.594,	.594,	.594}
DK	.626	EE	{.535,	.556,	.574}	{.554,	.564,	.573}	{.572,	.572,	.572}
HU	.623	HU	{.468,	.518,	.558}	{.509,	.532,	.553}	{.547,	.547,	.547}
EE	.589	PL	{.505,	.523,	.537}	{.523,	.531,	.538}	{.539,	.539,	.539}
PL	.567	DK	{.496,	.513,	.526}	{.513,	.521,	.528}	{.530,	.530,	.530}
SK	.563	SK	{.467,	.475,	.482}	{.476,	.480,	.484}	{.486,	.486,	.486}
RO	.562	LV	{.440,	.459,	.474}	{.458,	.466,	.474}	{.474,	.474,	.474}
LV	.500	RO	{.446,	.456,	.466}	{.457,	.462,	.467}	{.469,	.469,	.469}
DE	.488	CH	{.402,	.427,	.448}	{.424,	.435,	.446}	{.443,	.443,	.443}
CH	.474	LU	{.357,	.389,	.414}	{.382,	.397,	.410}	{.405,	.405,	.405}
BE	.466	DE	{.339,	.384,	.423}	{.371,	.394,	.414}	{.404,	.404,	.404}
LU	.464	IE	{.360,	.386,	.405}	{.382,	.393,	.403}	{.401,	.401,	.401}
FR	.450	FR	{.340,	.380,	.408}	{.371,	.389,	.404}	{.399,	.399,	.399}
IE	.425	BE	{.318,	.373,	.409}	{.359,	.382,	.401}	{.392,	.392,	.392}
CY	.415	CY	{.300,	.336,	.362}	{.327,	.343,	.357}	{.351,	.351,	.351}
NO	.393	NO	{.291,	.324,	.347}	{.316,	.330,	.343}	{.337,	.337,	.337}
UK	.324	UK	{.290,	.304,	.315}	{.302,	.309,	.314}	{.314,	.314,	.314}
SI	.268	SI	{.250,	.258,	.264}	{.257,	.261,	.264}	{.264,	.264,	.264}

As we can see, a crisp index score is achieved when using the IDEA-based CI model, which is easy for communication. In the FDEA-based CI model, fuzzy index scores are obtained based on different possibility levels of h. In practice, the given possibility degree by decision makers reflects their attitude on uncertainty. When h=1, the ordinal data are actually treated as numerical ones and the same index scores are obtained for each country, no matter whether the decision makers are in a pessimistic, indifferent, or optimistic consideration.

When the given value of h becomes lower, it means the decision makers are more cautious. As a consequence, a wider range of index scores will be derived. In such a way, the uncertainties associated with human thinking are effectively interpreted. Taking Belgium as an example, which was assigned the lowest value of 3 for this ordinal indicator among all the 28 European countries, it obtains an index score of 0.392 when h=1. That is, decision makers have no doubt about this value in representing the true performance of Belgium with respect to this indicator, which is half of the value of 6 and one third of 9. When h decreases to 0.5, this implies that decision makers are no longer fully sure about the relation between 3 and 6, and the other numbers. In other words, the value of 6 could be more (or less) than twice as large as the value of 3, and the most that can be judged is that the former one is preferred to or more important than the latter. As a result, an interval index score is obtained for Belgium, which is between 0.359 (pessimistic) and 0.401 (optimistic), with a medium value of 0.382 (indifferent). The widest interval is derived when h=0, which is {0.318, 0.373, 0.409}. Among all the 28 European countries, Sweden is the only non-dominated best performing country since it obtains the fuzzy index score of one at all possibility levels h.

Moreover, by comparing the alcohol performance index scores of the 28 European countries derived from these two models, we find that the IDEA-CI score is greater than the one from the FDEA-based CI model, even in the optimistic situation with the lowest possibility level of h. This can be partly explained by the fact that a relatively small value of ε is used in the IDEA-based CI model to reflect the minimum allowable gap between the two ranking positions in terms of the indicator value, which results in an extreme index score for each country. In other words, based on the same ε value, the index score from the FDEA-based CI model would not exceed the one from the IDEA-based CI model. Nevertheless, a high correlation coefficient (0.989) is deduced between the IDEA-CI score and the FDEA-CI score (taking h=0.5 and the indifferent situation as an example). This not only demonstrates the robustness of their ranking results, but also implies the reliability of using these two approaches for modeling qualitative data.

8.5 A Composite Road Safety Performance Index based on the Fuzzy MLDEA-based CI Model

To construct a composite road safety performance index, in addition to the three alcohol indicators considered in the above section, all the other indicators developed in Chapter 5 have to be modeled simultaneously. In doing so, we not only need to make a difference between quantitative and qualitative indicators, but the hierarchical structure of the indicators should also be taken into account. To this end, the models developed in this chapter have to be integrated into the MLDEA-based CI model proposed in Chapter 6. In this respect, a fuzzy MLDEA-based CI model seems to be the only option because the weights of the qualitative indicators in the IDEA-based CI model are not attainable, which however, are indispensable information for the multilayer model. Based on the FDEA-based CI model (8-9) and the MLDEA-based CI model (6-8), we obtain the fuzzy MLDEA-based CI model as follows:

$$
\begin{aligned}
CI_c = \max\ & \lambda_1 \sum_{f_1=1}^{s} \hat{u}_{f_1}\left(y_{f_1c} - (1-h)a_{f_1c}\right) + \lambda_2 \sum_{f_1=1}^{s} \hat{u}_{f_1}\left(y_{f_1c} + (1-h)a_{f_1c}\right) \\
s.t. \quad & \sum_{f_1=1}^{s} \hat{u}_{f_1}\left(y_{uf_1j} + (1-h)a_{fj}\right) \le 1, \quad j = 1,\cdots,n \\
& \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Big/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{p_{f_k}^{(k)}} \in \Theta, \quad f_k = 1,\cdots,s^{(k)}, \quad k = 1,\cdots,K-1 \\
& \hat{u}_{f_1} \ge \varepsilon, \quad f_1 = 1,\cdots,s
\end{aligned}
\qquad (8\text{-}10)
$$

Using the same data normalization procedure and weight restrictions as in Chapter 6, we now apply the model (8-10) to combine all the 33 indicators into a composite road safety performance index for the 28 European countries. Given the possibility level of h=0.5, the results are obtained in Table 8.5, together with the ones from Chapter 6, in which 32 quantitative indicators were combined based on the MLDEA-based CI model.

Table 8.5 Composite road safety performance index scores of 28 European countries based on the fuzzy MLDEA-based CI model and the MLDEA-based CI model

	FMLDEA-CI (h=0.5)			MLDEA-CI
NL	{0.99245,	0.99622,	1.00000}	1.00000
LU	{0.99238,	0.99619,	1.00000}	1.00000
CH	{0.99233,	0.99616,	1.00000}	1.00000
SE	{0.99228,	0.99614,	1.00000}	1.00000
DE	{0.99224,	0.99612,	1.00000}	1.00000
AT	{0.99114,	0.99557,	1.00000}	0.96509
UK	{0.98161,	0.98767,	0.99372}	0.97092
ES	{0.96400,	0.97081,	0.97762}	0.96139
EE	{0.95583,	0.96429,	0.97276}	0.92779
NO	{0.95056,	0.95664,	0.96273}	0.94099
FI	{0.94447,	0.95072,	0.95697}	0.90558
FR	{0.94336,	0.94740,	0.95144}	0.94450
CZ	{0.93308,	0.94072,	0.94835}	0.91156
IT	{0.93163,	0.93701,	0.94240}	0.96693
DK	{0.92669,	0.93263,	0.93857}	0.88404
SI	{0.92253,	0.93008,	0.93763}	0.91269
BE	{0.92102,	0.92504,	0.92905}	0.95820
IE	{0.91423,	0.92028,	0.92633}	0.90484
PT	{0.90986,	0.91819,	0.92652}	0.90347
SK	{0.90424,	0.91271,	0.92117}	0.87693
PL	{0.86891,	0.87491,	0.88090}	0.84510
BG	{0.85018,	0.85555,	0.86092}	0.86898
CY	{0.83444,	0.83910,	0.84377}	0.83373
LT	{0.81126,	0.81872,	0.82618}	0.80520
RO	{0.80449,	0.81277,	0.82105}	0.76630
LV	{0.78196,	0.78942,	0.79689}	0.74608
EL	{0.77192,	0.77943,	0.78694}	0.74900
HU	{0.75244,	0.75990,	0.76737}	0.75146

Due to the inclusion of one more, ordinal, indicator in the composite index construction, a fuzzy rather than crisp index score is obtained for each country based on the fuzzy MLDEA-based CI model. In general, the two sets of index scores are highly correlated with a correlation coefficient of 0.966 when the FMLDEA-CI score under the indifferent situation is considered. This is mainly attributed to the use of the multilayer model, which reduces the sensitivity of the final index score when adding new indicators into the model (or deleting existing indicators from the model). However, differences can still be found. For instance, Austria becomes another (fuzzy) best-performing country apart from the Netherlands, Luxembourg, Switzerland, Sweden, and Germany, since a relatively high alcohol enforcement indicator value (9) is assigned to it. At the same time, the ranking of Belgium has a considerable decline due to the fact that it has the lowest score (3) with respect to its drink driving enforcement. In addition, Hungary becomes the worst-performing country even under the most optimistic consideration, because it obtained a relatively lower indicator value (5) compared with countries such as Greece and Latvia, which were both assigned a score of 7 for this indicator.

8.6 Conclusion

In this chapter, we investigated two approaches within the DEA framework for modeling qualitative (or ordinal) data in the context of composite index construction. They are imprecise DEA and fuzzy DEA, respectively. Taking their principle for reference, we proposed two new models – the IDEA-based CI model and the FDEA-based CI model – for road safety performance evaluation. Based on three alcohol indicators (two quantitative and one qualitative), we successfully obtained the composite alcohol performance index scores for the 28 European countries. The analysis of the results showed that the crisp index score achieved by the IDEA-based CI model is easy for interpretation and use, while in the FDEA-based CI model, fuzzy index scores obtained based on different possibility levels are powerful in capturing the uncertainties associated with human thinking. The high similarity of the ranking result based on these two models verifies its robustness and implies the reliability of using these two approaches for modeling qualitative data. Furthermore, by integrating fuzzy logic into the MLDEA-based CI model proposed in Chapter 6, we obtained the fuzzy MLDEA-based CI model, which can not only reflect the hierarchical structure of the indicators, but also make the differentiation between quantitative and qualitative indicators possible. The model was therefore capable of combining all the 33 hierarchical indicators (with both quantitative and qualitative data) into a composite road safety performance index.

Chapter 9 Final Conclusions and Future Research

This chapter summarizes the main contributions and findings from this dissertation research. Furthermore, directions for future research in relation to this topic are provided.

9.1 General Conclusions

Inter-national benchmarking of road safety performance and development is considered as a promising step to improve a country's road safety level and has currently been widely advocated by most countries and international bodies. Having recognized the complex character of the road safety phenomenon, different aspects of the road safety management and improvement process have been identified, and a large number of road safety indicators have been developed within each aspect and increasingly used as a supportive instrument for inter-national comparisons and monitoring of road safety progress. In this dissertation research, we aimed to implement meaningful benchmarking on the road safety product (using risk indicators) and the road safety programme (using safety performance indicators) for 28 European countries, with the purpose of better understanding each country's relative road safety situation, identifying the areas of underperformance per country, and moreover, enabling policymakers to learn from those better-performing countries as a basis for developing their own road safety policy. To achieve these objectives, the technique of data envelopment analysis (DEA) – one of the powerful benchmarking tools currently receiving considerable attention in the operations research and economics field – was introduced to the road safety domain, and its various extensions were comprehensively investigated to handle some of the model limitations and to answer the specific research questions that have not yet been properly addressed in current road safety benchmarking studies. The main contributions and findings from this dissertation research are summarized as follows:

i) Development of a DEA-based road safety model and its extensions for road safety risk evaluation

Road safety researchers have traditionally evaluated the road safety performance of a country by comparing its road safety risk indicators with those of other countries. However, in computing the level of risk for a country, which is commonly defined as the ratio of road safety outcomes such as the number of road fatalities and some measure of exposure, different exposure information can be used (e.g., the population size, the number of registered vehicles, and the distance travelled), and different evaluation results or ranking positions are normally obtained based on different risk indicators. So far, there is no consensus about which one is the most appropriate indication, and research on their combination is also quasi non-existing.

In this research, using the principle of DEA for reference, which is to measure the relative efficiency of a set of decision making units (DMUs) on the basis of multiple inputs and multiple outputs, we developed a DEA-based road safety model (DEA-RS) to evaluate the overall road safety performance of 28 European countries by considering the three main risk indicators (i.e., the number of fatalities per million inhabitants, the number of fatalities per 10 billion passenger-kilometres travelled, and the number of fatalities per million passenger cars) simultaneously. Using the model linking input (three measures of exposure to risk) and output (the number of fatalities), an overall road safety efficiency score was obtained for each country indicating its level of efficient transformation of input or exposure into output or road safety fatalities, and the ranking of these countries was deduced by computing their cross-efficiency. The result gave us a global view on the country's road safety performance by taking all three aspects of exposure into account, and yet it was not the simple average of those three rankings.

Moreover, after performing clustering analysis to group countries with inherent similarity in their practices, we applied a categorical DEA-RS model to identify best-performing and underperforming countries in each group, and to indicate appropriate reference sets or benchmarks for those underperforming ones. More importantly, the extent to which each reference set could be learned from was specified, and several practical targets on the number of road fatalities were given for each underperforming country. All these enable policymakers to recognize the gap with those best-performing countries within each group and to further develop their own road safety policy.

One additional advantage of using the DEA-RS model for road safety risk evaluation is that the model can be easily extended when other inputs and/or outputs are taken into account. In Chapter 4, we explored the impact of including serious injuries in addition to the fatalities as another road safety outcome for road safety risk evaluation. The DEA-RS model was successfully applied by imposing additional weight restrictions to indicate the relationship between road fatalities and serious injuries. We are therefore inspired to apply this model to a more complete road safety product benchmarking practice in the future when the data on for example the number of crashes, the degree of property damage, and the number of slight injuries are ready to use.

ii) Adoption of the Malmquist productivity index to evaluate road safety performance change over time

Apart from comparing the road safety risk indicators at one specific point of time, the analysis of the performance change of countries over time has also been conducted in the road safety product benchmarking. Traditionally, the percentage change in the number of people killed on the road is the main indicator used to compare the development of road safety between countries with a higher reduction in road fatalities indicating a better rank. However, simply considering the reduction in the final outcome may not correctly reflect the real improvement in road safety because the transport circumstances of a country underlying the final outcome also change every year.

In this research, we presented a new way to assess the road safety performance change of countries over time, which was by applying the Malmquist productivity index. The index has the capability to capture the dynamic road safety development in each country by not only focusing on the evolution of road safety final outcomes within a given period, but also taking the changes in exposure in the same period into account.

In the application, using the information on the three measures of exposure, i.e., the number of inhabitants, passenger-kilometres travelled and passenger cars on the one hand, and the number of fatalities in road transport on the other hand, the Malmquist productivity index based on the DEA-RS model has proven valuable in measuring the extent to which the 28 European countries have improved their 'productivity' on road safety over the period 2000-2009. It is believed to have provided more objective results than the ones based on the traditional indicator that only measures the percentage change in road fatalities. Moreover, the decomposition of the index into efficiency change and technical change further provides policymakers with valuable information to gain a clear understanding on whether the improvement in road safety of each country was attained through country-specific progress relative to the others, or just through an overall improvement in the technological environment.

iii) Realization of a multiple layer DEA-based composite index model for hierarchical structure assessment

With the continuously growing awareness of the complex character of the road safety phenomenon, it has been widely acknowledged that the traditional way in assessing a country's road safety situation that only concentrates on the road safety final outcomes is insufficient in explaining more detailed aspects of crash causation and injury prevention. Over the last decade, more and more policy attention has been paid to the underlying risk factors influencing safety or, at least, those factors policymakers are able to affect or control. In this research, a comprehensive set of safety performance indicators (SPIs), which are viewed as intermediate outcomes linking safety countermeasures with final outcomes, was developed based on the identification of six leading road safety risk factors (i.e., alcohol, speed, protective systems, vehicle, road, and emergency medical services) within the three main road transport components (i.e., road user, vehicle and infrastructure), which provides the basis for the second type of benchmarking practice, i.e., road safety programme benchmarking.

Since a number of SPIs are considered for a particular risk factor in this research, and they are further linked to one another constituting a layered hierarchy, a simple comparison per indicator may only show a small piece of the road safety picture, and it can be misleading since different countries may operate in different circumstances with different focal points. Consequently, a composite road safety performance index, which combines individual indicator values into one single score, is valuable to be computed for the sake of meaningful benchmarking. In doing so, the hierarchical structure of the SPIs is also worthwhile to be reflected. However, it is prone to be ignored in the current index research due in part to the limitation of traditional weighting and aggregation techniques in reflecting this kind of hierarchical structures.

In this dissertation research, we investigated the use of DEA to develop a composite road safety performance index for cross-country comparison due to its prominent advantages over other traditional methods. Moreover, we further explored the incorporation of a layered hierarchy in the DEA framework, and proposed a multiple layer DEA-based composite index model (MLDEA-CI) for hierarchical structure assessment.

In the application, we used the proposed MLDEA-CI model to determine the most optimal road safety performance index score for each of the 28 European countries by combining all the 32 hierarchical SPIs. Countries were then classified into two groups: best-performing and underperforming, and the ranking of the countries was deduced by computing their cross-index score. A clear link with the overall road safety risk from the view of the final outcome level was verified, which in turn justified the use of the proposed multilayer model for composite index construction. Furthermore, by taking the characteristics of each country in the data set into account, country-specific benchmarks were identified for the underperforming countries. More importantly, by analyzing the indicator weights allocated in each layer of the hierarchy, useful insight in the areas of underperformance in each country was gained enabling policymakers to prioritize their actions to improve the level of road safety.

iv) Consideration of data issues in the composite road safety performance index construction

Coupled with the proliferation of SPIs, some practical issues related to data also inevitably emerge in the development of a composite road safety performance index, two of which are qualitative indicators and missing values. Specifically, obtainment of measurable and quantitative indicators is commonly a prerequisite of any index research. This, however, becomes more and more difficult to be guaranteed since the natural uncertainty of reality often leads up to imprecision and vagueness inherent in the information that can only be represented by means of qualitative indicators. Simply treating them as quantitative ones could thereby result in wrong conclusions. Moreover, an extension of the data set used for road safety index research raises the issue of missing values, which to a great extent restricts researchers from performing classical analyses as complete data matrices are usually required. Consequently, how to effectively handle these data problems directly affects the result of the road safety index research and the success of benchmarking practices as well.

In Chapter 7, we explored the influence of the existence of missing data in the data set on the final index score of the 28 European countries. In doing so, missing data were firstly replaced by approximations in the form of intervals deduced from multiple imputation in which the true values are believed to lie. Thus, the complete data set consisted of both observed data and imputed data intervals. Subsequently, an interval MLDEA-based CI model was applied and resulted for each country in an upper and a lower bound of its index score corresponding to its most favorable and unfavorable option, respectively. The interval instead of the precise index score for each country highlights the underlying imperfect nature of the indicator data, and provides us with a more credible representation of a country's overall road safety performance. Based on the interval index scores, countries could be further classified into 'benchmark countries', 'potential benchmark countries', and 'countries open to improvement'.

In Chapter 8, we investigated two approaches within the DEA framework for modeling qualitative (or ordinal) data in the context of composite index construction, being the imprecise DEA-based CI model and the fuzzy DEA-based CI model. The analysis of the results shows that the crisp index score achieved by the IDEA-based CI model is easy for interpretation and use, while in the FDEA-based CI model, fuzzy index scores obtained based on different possibility level are powerful in capturing the uncertainties associated with human thinking. The high similarity of the ranking result based on these two models verifies its robustness and implies the reliability of using these two approaches for modeling qualitative data. Furthermore, a fuzzy MLDEA-based CI model was realized to construct a composite road safety performance index based on both quantitative and qualitative indicators.

To conclude, we have identified in this dissertation the main research issues with respect to the road safety product and programme benchmarking, respectively, and have developed corresponding approaches to deal with these issues. This research has contributed to the literature on using the technique of DEA and its various extensions to perform meaningful inter-national benchmarking of road safety performance and development based on different types of road safety indicators. Although it is mathematical in nature, we should say that the theory behind this technique is straightforward and it is currently ready for implementation at the practical level. In addition, the added value for road safety policymakers lies in the development of a composite road safety performance index as a facility for each country to assess its own overall road safety performance, and moreover, in the formulation of policy recommendations with respect to both target setting and action prioritizing to improve the level of road safety in their country. All this, in turn, forms a strong foundation for future research, which will be elaborated in the next section.

9.2 Topics for future research

Road safety is an important policy area that can benefit from the implementation of various inter-national benchmarking practices. In this respect, an appropriate selection of road safety indicators, a harmonized data collection procedure, and a scientifically sound methodology are the fundamental conditions of making meaningful comparisons between countries, and also the key to designing more effective safety policies. However, given the fact that road safety research towards a thorough understanding of this complex problem is still an ongoing process, the degree of maturity of the developed road safety indicators and the availability and quality of indicator data are still open for discussion, and a number of methodological challenges remain for future research.

Currently, most of the road safety product benchmarking practices focus entirely on fatalities, which however, represent only the ‘tip of the iceberg’ of the road crash problem and could lead up to an overestimation of this aspect. Therefore, a first extension of the research in this respect is to take a larger picture of the impact of road crashes into account, such as the number of crashes and the range of injury severities. In Chapter 4, an initial attempt of including the number of serious injuries as an additional indicator for road safety product benchmarking was carried out. However, it is not yet a mature indicator due to large differences in reporting practices in different countries. Consequently, international cooperation in terms of crash/injury data collection and harmonization is sorely required, which will be beneficial to all bodies that are concerned with road safety management. Second, from the view of exposure selection, the total number of motor vehicles and the number of motor vehicle kilometres travelled rather than the number of passenger cars and the passenger-kilometres travelled are much more preferred to use as a true representation of the level of a country's motorization, as soon as the data on these measures are available. Moreover, there are still some other variables that can be used as measures of exposure under specific circumstances, such as road length, fuel consumption, the number of driving license holders, and so on. They are all potential candidates to be applied for road safety risk evaluation in the future. In addition, analysis at the disaggregated level of both road safety outcomes and exposure (such as based on age group, gender, person class, transport mode, and area) is also valuable as disaggregated data allow the examination of unique interactions in a way that aggregated data cannot.

With respect to the road safety programme benchmarking, indicators developed for most of the risk factors in this research are extensive and comprehensive based on our current knowledge. However, reliable and comparable indicator data, especially concerning alcohol, speed, and emergency medical services, are still lacking to some extent. Regarding the factor of road, only limited and proxy indicators and data are currently available for benchmarking purposes. Knowledge on the quantitative relations between the road network, road design elements and road safety therefore needs further exploration, and a variety of appropriate indicators corresponding to this aspect call for different kinds of development efforts relating to concepts, methodologies, and data collection procedures. Moreover, other risk factors that have a strong relationship with road safety or a large contribution to road crashes or casualties, such as inattentive driving as a result of mobile technology, could also be incorporated in the future and corresponding indicators developed and refined. Furthermore, by collecting the safety performance data at regular intervals, systematic country comparison over time could be conducted so as to evaluate the results of policy interventions and to monitor the progress in road safety performance.

In addition to the road safety product and programme benchmarking investigated in this research, other aspects of the road safety management and improvement process, such as road safety strategic and organizational benchmarking, which are used to compare national road safety strategies, resources, management and the organizational framework, are also desirable to be implemented in the future, and the interrelation between different benchmarking practices can be studied in detail.

From the view of the methodology used in this research, the technique of DEA has proven to be a powerful benchmarking tool for providing interesting insights and valuable recommendations in both road safety product and programme benchmarking studies, and its various extensions, including the DEA-based road safety model (DEA-RS), the categorical DEA-RS model, the multiple layer DEA-based composite index model (MLDEA-CI), the interval MLDEA-CI model, the imprecise DEA-CI model, and the fuzzy (ML)DEA-CI model, have been successfully developed and applied to handle the specific research issues associated with the indicators and the data. In the future, due to the continuous development and update of appropriate indicators for different benchmarking purposes, some new methodological challenges will probably appear (such as the treatment of negative indicator values, undesirable factors, and non-discretionary (or environmental) variables), which have to be tackled accordingly and the solutions should be able to be correctly integrated with the existing models. Moreover, as indicated in the beginning of this thesis, DEA is regarded as a body of concepts and methodologies that has evolved since the seminal work of Charnes, Cooper & Rhodes (1978). Therefore, apart from the CCR model introduced and applied in this research, the added value of using a large number of other DEA models (such as the BCC model, the additive model, the slacks-based measure of efficiency, and the multiplicative model) can also be investigated in road safety benchmarking studies in the future. Furthermore, since the results obtained from DEA are sensitive to country selection, indicator specification, data quality and chosen weight restrictions, more research attention should be paid to the sensitivity and stability analysis of DEA [Simar & Wilson, 2000]. In this respect, statistical tests for DEA [Jenkins & Anderson, 2003], bootstrapping in DEA [Tortosa-Ausina et al., 2008], and stochastic DEA [Huang & Li, 2001] are all worthwhile to be explored. Moreover, it would be interesting to perform in the future an empirical investigation on whether underperforming countries would choose the specific benchmarks indicated in this research as it will help in determining the validity of the methodology.

Inter-national benchmarking of road safety performance and development has and will continue to play an important role in improving a country's road safety level. Nevertheless, from the road safety policy point of view, we should always keep in mind that benchmarking does not represent the end of the process, but is an ongoing diagnostic management tool requiring effective strategies, sufficient allocation of resources, successful implementation, and persistent monitoring and evaluation in order to achieve continuous improvement over time.

References

Adolphson, D.L., Cornia, G.C. & Walters, L.C., (1991). A Unified Framework for Classifying DEA Models, In: Operational Research '90, Pergamon Press, New York, pp. 647-657.

Aigner, D.J., Lovell, C.A.K. & Schmidt, P., (1977). Formulation and estimation of stochastic frontier production functions, Journal of Econometrics, Vol. 6, pp. 21-37.

Allen, R., Athanassopoulos, A., Dyson, R.G. & Thanassoulis, E., (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions, Annals of Operations Research, Vol. 73, pp. 13-34.

Allison, P.D., (2001). Missing Data Thousand Oaks, Sage Publications, California.

Allsop, R.E., Sze, N.N. & Wong, S.C., (2011). An update on the association between setting quantified road safety targets and road fatality reduction, Accident Analysis & Prevention, Vol. 43, No. 3, pp. 1279-1283.

Al-Haji, G., (2007). Road Safety Development Index (RSDI): Theory, Philosophy and Practice. Department of Science and Technology. PhD Dissertation, Campus Norrköping, Linköping University, Norrköping.

American Productivity and Quality Center, (1993). The Benchmarking Management Guide, Productivity Press, Inc., Portland.

Amirteimoori, A. & Kordrostami, S., (2005). Allocating fixed costs and target setting: A DEA-based approach, Applied Mathematics and Computation, Vol. 171, pp. 136-151.

Andersen, B., (1995), The Results of Benchmarking and a Benchmarking Process Model, Ph.D. Dissertation, Norwegian Institute of Technology, Trondheim, Berlin.

Andersen, B. & Pettersen, P., (1996). The Benchmarking Handbook: Step-by-Step Instructions. Chapman and Hall, London, pp. 13-15.

Anderson, T.R., Hollingsworth, K. & Inman, L., (2002). The fixed weighting nature of a cross-evaluation model, Journal of Productivity Analysis, Vol. 17, pp. 249-255.

Bandura, R., (2008). A Survey of Composite Indices Measuring Country Performance: 2008 Update, United Nations Development Programme - Office of Development Studies.

Banker, R.D., Charnes, A. & Cooper, W.W., (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management Science, Vol. 30, pp. 1078-1092.

Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M. & Zhu, J., (2004). Returns to scale in different DEA models, European Journal of Operational Research, Vol. 154, pp. 345-362.

Bao, Q., Ruan, D., Shen, Y., Hermans, E. & Janssens, D., (2012). Improved hierarchical fuzzy TOPSIS for road safety performance evaluation, Knowledge-Based Systems, doi: 10.1016/j.knosys.2011.08.014.

Barnett, V. & Lewis, T., (1994). Outliers in Statistical Data, John Wiley & Sons, Chichester.

Ben-Gal, I., (2005). Outlier Detection, In: Maimon, O. & Rockach, L. (Eds.), Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Kluwer Academic Publishers, Boston.

Bessent, A., Bessent, W., Kennington, J. & Reagan, B., (1982). An application of mathematical programming to assess productivity in the Houston independent school district, Management Science, Vol. 28, No. 2, pp. 82-107.

Bester, C.J., (2001). Explaining national road fatalities, Accident Analysis & Prevention, Vol. 33, No. 5, pp. 663-672.

Bhutta, K.S. & Huq, F., (1999). Benchmarking-Best practices: An integrated approach, Benchmarking: An International Journal, Vol. 6 No. 3, pp. 254-268.

Bogetoft, P. & Otto, L., (2011). Benchmarking with DEA, SFA, and R, International Series in Operations Research & Management Science, Vol. 157, Springer.

Boussofiane, A., Dyson, R.G. & Thanassoulis, E., (1991). Applied data envelopment analysis, European Journal of Operational Research, Vol. 52, pp. 1-15.

Bowie, N.N. & Waltz, M., (1994). Data analysis of the speed-related crash issue, Auto and Traffic Safety, Vol. 1, No. 2, pp.31-38.

Broughton, J., Amoros, E., Bos, N., Evgenikos, P., Hoeglinger, S., Holló, P., Pérez, C. & Tecl, J., (2008). Final Report on Task 1.5, Deliverable D1.15 of the EU FP6 project SafetyNet.

Camp, R.C., (1989). Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance, ASQ Quality Press, Milwaukee.

Camp, R.C., (1995). Business Process Benchmarking: Finding and Implementing Best Practices, ASQ Quality Press, Milwaukee.

Caussinus, H. & Roiz, A., (1990). Interesting projections of multidimensional data by means of generalized component analysis, In Compstat 90, Physica-Verlag, Heidelberg, pp. 121-126.

Caves, D.W., Christensen, L.R. & Diewert, W.E., (1982). The economic theory of index numbers and the measurement of input, output, and productivity, Econometric, Vol. 50, No. 6, pp. 1393-1414.

Charnes, A., Cooper, W.W. & Rhodes, E., (1978). Measuring the efficiency of decision making units, European Journal of Operational Research, Vol. 2, pp. 429-444.

Charnes, A. & Cooper, W.W., (1984). The non-Archimedean CCR ratio for efficiency analysis: A rejoinder to Boyd and Fare, European Journal of Operational Research, Vol. 15, No. 3, pp. 333-334.

Charnes, A., Cooper, W.W., Huang, Z. & Sun, D., (1990). Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks, Journal of Econometrics, Vol. 46, pp. 73-91.

Charnes, A., Cooper, W.W., Lewin, A.L. & Seiford, L.M., (1994). DEA: Theory, Methodology, and Application, Kluwer Academic Publishers, Boston.

Chen, Y. & Ali, A.I., (2004). DEA Malmquist productivity measure: New insights with an application to computer industry, European Journal of Operational Research, Vol. 159, pp. 239-249.

Chen, J. & Shao, J., (2000). Nearest neighbour imputation for survey data, Journal of Official Statistics, Vol. 16, No. 2, pp. 113-131.

Cheng, M., Tsai, M. & Sutan, W., (2009). Benchmarking-based process reengineering for construction management, Automation in Construction, Vol. 18, No. 5, pp. 605-623.

Cherchye, L., Moesen, W., Rogge, N. & van Puyenbroeck, T., (2007a). An introduction to 'benefit of the doubt' composite indicators, Social Indicators Research, Vol. 82, pp. 111-145.

Cherchye, L., Lovell, C.A.K., Moesen, W., & van Puyenbroeck, T., (2007b). One market, one number? A composite indicator assessment of EU internal market dynamics, European Economic Review, Vol. 51, pp. 749-779.

Cherchye, L., Moesen, W., Rogge, N., van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R. & Tarantola, S., (2008). Creating composite indicators with DEA and robustness analysis: The case of the technology achievement index, Journal of the Operational Research Society, Vol. 59, pp. 239-251.

Cherchye, L., Moesen, W., Rogge, N. & Van Puyenbroeck, T. (2011). Constructing a knowledge economy composite indicator with imprecise data: A proposal, Expert Systems with Applications, Vol. 38, No. 9, 10940-10949.

Chung, W., (2011). Review of building energy-use performance benchmarking methodologies, Applied Energy, Vol. 88, No. 5, pp. 1470-1479.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J. & Battese, G.E., (2005). An Introduction to Efficiency and Productivity Analysis, 2nd Edition. Springer.

Cook, W.D. & Kress, M., (1990). A data envelopment model for aggregating preference rankings, Management Science, Vol. 36, No. 11, pp. 1302-1310.

Cook, W.D., Kress, M. & Seiford, L.M., (1993). On the use of ordinal data in data envelopment analysis, Journal of the Operational Research Society, Vol. 44, pp. 133-140.

Cook, W.D., Kress, M. & Seiford, L.M., (1996). Data envelopment analysis in the presence of both quantitative and qualitative factors, Journal of the Operational Research Society, Vol. 47, pp. 945-953.

Cook, W.D. & Zhu, J., (2006). Rank order data in DEA: A general framework, European Journal of Operational Research, Vol. 174, pp. 1021-1038.

Cook, W.D. & Zhu, J., (2007). Rank Order Data in DEA, In: Zhu, J. & Cook, W.D. (Eds.), Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis, Springer, New York, pp. 103-122.

Cook, W.D. & Seiford, L.M., (2009). Data envelopment analysis (DEA) – Thirty years on, European Journal of Operational Research, Vol. 192, pp. 1-17.

Cooper, W.W., Park, K.S. & Yu, G., (1999). IDEA and AR-IDEA: Models for dealing with imprecise data in DEA, Management Science, Vol. 45, pp. 597-607.

Cooper, W.W., Seiford, L.M. & Tone, K., (2000), DEA: A Comprehensive Text with Models, Application, References and DEA Solver Software, Kluwer Academic Publishers, Boston.

Cooper, W.W., Park, K.S. & Yu, G., (2002). An illustrative application of IDEA (Imprecise Data Envelopment Analysis) to a Korean mobile telecommunication company, Operations Research, Vol. 49, No. 6, pp. 807-820.

Cooper, W.W., Seiford, L.M. & Zhu, J., (2004). Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston.

Davies, L. & Gather, U., (1993). The identification of multiple outliers, Journal of the American Statistical Association, Vol. 88, No. 423, pp. 782-792.

Dempster, A.P. & Rubin, D.B., (1983). Introduction. In: Madow, W.G., Olkin I. & Rubin D.B. (Eds.), Incomplete Data in Sample Surveys, Vol. 2: Theory and Bibliography, Academic Press, New York, pp.3-10.

Despotis, D.K. & Smirlis, Y.G., (2002). Data envelopment with imprecise data, European Journal of Operational Research, Vol. 140, pp. 24-36.

Despotis, D.K., (2005). Measuring human development via data envelopment analysis: The case of Asia and the Pacific, Omega, International Journal of Management Science, Vol. 33, pp. 385-390.

Doyle, J. & Green, R., (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses, Journal of the Operational Research Society, Vol. 45, pp. 567-578.

Dyson, R.G., & Thanassoulis, E., (1988). Reducing weight flexibility in data envelopment analysis, Journal of the Operational Research Society, Vol. 39, No. 6, pp. 563-576.

Eksler, V., (2009). Road Mortality in Europe: A Regional Approach, Ph.D. Dissertation, UVSQ, Saint-Quentin-en-Yvelines.

Eksler, V., Popolizio, M., & Allsop, R., (2009). How far from Zero? Benchmarking of Road Safety Performance in the Nordic Countries, European Transport Safety Council, Brussels.

Elmuti, D., Kathawala, Y. & Lloyed, S., (1997). The benchmarking process: Assessing its value and limitations, Industrial Management, Vol. 39, No. 4, pp. 12–19.

Elvik, R., (2001). Quantified Road Safety Targets: An Assessment of Evaluation Methodology, TØI Report 539/2001, Institute of Transport Economics, Oslo.

Elvik, R. (2003). An Overview of Target-setting in Europe: Best in Europe Conference on Targeted Road Safety Programmes in the EU, European Transport Safety Council, Brussels.

Elvik, R. & Vaa, T., (2004). Handbook of Road Safety Measures, Elsevier Ltd, Oxford.

Elvik, R., (2005). Speed and road safety: synthesis of evidence from evaluation studies, 84th annual meeting of the Transportation Research Board, Washington.

El-Mahgary, S. & Lahdelma, R., (1995). Data envelopment analysis: Visualizing the results, European Journal of Operational Research, Vol. 85, pp. 700-710.

Emrouznejad, A., Parker, B.R. & Tavares, G., (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA, Journal of Socio-Economics Planning Science, Vol. 42, No. 3, pp. 151-157.

Entani, T., Maeda, Y. & Tanaka, H., (2002). Dual models of interval DEA and its extension to interval data. European Journal of Operational Research, Vol. 136, pp. 32-45.

Ertay, T. & Ruan, D., (2005). Data envelopment analysis based decision model for optimal operator allocation in CMS, European Journal of Operational Research, Vol. 164, pp. 800-810.

European Commission (EC), (2001). White Paper: European Transport Policy for 2010: Time to Decide, Commission of the European Communities, Brussels.

European Commission (EC), (2003). European Road Safety Action Programme - Halving the Number of Road Accident Victims in the European Union by 2010: A Shared Responsibility, Commission of the European Communities, Brussels.

European Commission (EC), (2009). EC Public consultation on the European Road Safety Action Programme 2011-2020, Commission of the European Communities, Brussels. http://ec.europa.eu/transport/road_safety/take-part/public-consultations/action_en.htm.

European Commission (EC), (2010a). Towards a European Road Safety Area: Policy Orientations on Road Safety 2011-2020. COM (2010) 389 final, Commission of the European Communities, Brussels.

European Commission (EC), (2010b). EU Energy and Transport in Figures 2010, Commission of the European Communities, Brussels.

European Commission (EC), (2011a). EU Energy and Transport in Figures 2011. Commission of the European Communities, Brussels.

European Commission (EC), (2011b). http://ec.europa.eu/transport/road_safety/specialist/statistics/trends/index_en.htm.

European Conference of Ministers of Transport (ECMT), (1998). Efficient Transport for Europe: Policies for Internalisation of External Costs, ECMT, Paris.

European Road Safety Observatory (ERSO), (2006). Quantitative Road Safety Targets. http://www.erso.eu.

European Transport Safety Council (ETSC), (1999). Transport Safety Visions, Targets, and Stragegies: Beyond 2000, ETSC, Brussels.

European Transport Safety Council (ETSC), (2001). Transport Safety Performance Indicators, ETSC, Brussels.

European Transport Safety Council (ETSC), (2003). Assessing Risk and Setting Targets in Transport Safety Programmes, ETSC, Brussels.

European Transport Safety Council (ETSC), (2005). SECBelt Report: The Safety of Vulnerable Road Users in the Southern, Eastern and Central European Countries, ETSC, Brussels.

European Transport Safety Council (ETSC), (2007). Social and Economic Consequences of Road Traffic Injury in Europe, ETSC, Brussels.

European Transport Safety Council (ETSC), (2008). Road Safety as a Right and Responsibility for All: A Blueprint for the EU's 4th Road Safety Action Programme 2010-2020, ETSC, Brussels.

European Transport Safety Council (ETSC), (2009). 2010 on the Horizon, 3rd Road Safety PIN Report, ETSC, Brussels.

European Transport Safety Council (ETSC), (2010). Road Safety Target in Sight: Making up for Lost Time, 4th Road Safety PIN Report, ETSC, Brussels.

European Transport Safety Council (ETSC), (2011). 2010 Road Safety Target Outcome: 100,000 Fewer Deaths since 2001, 5th Road Safety PIN Report, ETSC, Brussels.

European Union Road Federation (ERF), (2010). 2010 European Road Statistics, ERF. www.erf.be.

Färe, R., Grosskopf, S., Lindgren, B. & Roos, P., (1992). Productivity change in Swedish pharmacies 1980-1989: A nonparametric Malmquist approach, Journal of Productivity Analysis, Vol. 3, pp. 85-102.

Färe, R., Grosskopf, S. & Hernández-Sancho, F., (2004). Environmental performance: An index number approach, Resource and Energy Economics, Vol. 26, pp. 343-352.

Farhangfar, A., Kurgan, L. & Pedrycz, W., (2004). Experimental analysis of methods for handling missing values in databases. In Proceedings of Intelligent Computing: Theory and Application II Conference, Orlando.

Farrell, M.J., (1957). The measurement of productive efficiency, Journal of the Royal Statistical Society, Vol. 120, No. 3, pp. 253-290.

Freudenberg, M., (2003). Composite Indicators of Country Performance: A Critical Assessment, STI working paper 2003/16, Organization for Economic Co-operation and Development, Paris.

Forum of European Road Safety Research Institutes & European Conference of Transport Research Institutes (FERSI/ECTRI), (2009). The Sustainable Safety Approach to Road Transport and Mobility, FERSI/ECTRI. http://www.ectri.org/Publications/Strategic-Documents.htm.

Garlick, S. & Pryor, G., (2004). Benchmarking the University: Learning about Improvement, Regional Knowledge Works, Department of Education Science and Training, Canberra.

Gaygısız, E., (2010). Cultural values and governance quality as correlates of road traffic fatalities: A nation level analysis, Accident Analysis & Prevention, Vol. 42, pp. 1894-1901.

Geraedts, M. & Selbmann, H., (2004). Benchmarking the management of projects: A review of current thinking, International Journal of Project Management, Vol. 22, No. 4, pp. 301-307.

Gitelman, V., Auerbach, K., Doveh, E., Avitzour, M. & Hakkert, S., (2008). Safety Performance Indicators for Trauma Management: Theory Update, Deliverable D3.11b EU FP6 Project SafetyNet.

Gitelman, V., Doveh, E. & Hakkert, S., (2010). Designing a composite indicator for road safety, Safety Science, Vol. 48, pp. 1212-1224.

Golany, B. & Roll, Y., (1989). An application procedure for DEA, Omega, International Journal of Management Science, Vol. 17, No. 3, pp. 237-250.

Green, M. & Senders J., (2004). Human Error in Road Accidents, Visual Expert, Canada. http://www.visualexpert.com/Resources/roadaccidents.html.

Greer, M.R., (2008). Nothing focuses the mind on productivity quite like the fear of liquidation: Changes in airline productivity in the United States, 2000-2004, Transportation Research Part A, Vol. 42, pp. 414-426.

Grzymala-Busse, J.W., (2008). Three approaches to missing attribute values: A rough set perspective, Studies in Computational Intelligence, Vol. 118, pp. 139-152.

Guo, P. & Tanaka, H., (2001). Fuzzy DEA: A perceptual evaluation method, Fuzzy Sets and Systems, Vol. 119, pp. 149-160.

Guo, P., (2009). Fuzzy data envelopment analysis and its application to location problems, Information Sciences, Vol. 179, pp. 820-829.

Guria, J., (2009). Required Expenditure. Road Safety Improvement in Low and Middle Income Countries. Addendum: Revised Estimates of Fatalities and Serious Injuries and Related Costs. Report to the World Bank Global Road Safety Facility, New Zealand Institute of Economic Research, Wellington.

Hadi, A.S., (1992). Identifying multiple outliers in multivariate data, Journal of the Royal Statistical Society, Series B, Vol. 54, pp. 761-771.

Hakkert, A.S. & Braimaister, L., (2002). The Uses of Exposure and Risk in Road Safety Studies, R-2002-12, SWOV Institute for Road Safety Research, Leidschendam.

Hakkert, A.S., Gitelman, V. & Vis, M.A. (Eds.), (2007a). Road Safety Performance Indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.

Hakkert, A.S. & Gitelman, V. (Eds.), (2007b). Road Safety Performance Indicators: Manual. Deliverable D3.8 of the EU FP6 project SafetyNet.

Hawkins, D., (1980). Identification of Outliers, Chapman and Hall, London.

Hayashi, F., (2000). Econometrics, Princeton University Press, Princeton.

Hermans, E., Van den Bossche, F. & Wets, G., (2008). Combining road safety information in a performance index, Accident Analysis and Prevention, Vol. 40, No. 4, pp. 1337-1344.

Hermans, E., (2009a). A Methodology for Developing a Composite Road Safety Performance Index for Cross-country Comparison, PhD Dissertation, Hasselt University, Hasselt.

Hermans, E., Brijs, T., Wets, G. & Vanhoof, K., (2009b). Benchmarking road safety: Lessons to learn from a data envelopment analysis, Accident Analysis and Prevention, Vol. 41, No. 1, pp. 174-182.

Hougaard, J.L., (2005). A simple approximation of productivity scores of fuzzy production plans, Fuzzy Sets and Systems, Vol. 152, No. 3, pp. 455-465.

Howell, D.C., (2008). The Analysis of Missing Data, In: Outhwaite, W. & Turner, S. (Eds.), Handbook of Social Science Methodology, Sage, London.

Huang, Z., Li, S.X., (2001). Stochastic DEA models with different types of input-output disturbances, Journal of Productivity Analysis, Vol. 15, pp. 95-113.

Hussain, I.M. & Redmond. A.D., (1994). Are pre-hospital deaths from accidental injury preventable? British Medical Journal, Vol. 308, pp. 1077-1080.

International Organization for Standardization (ISO), (2008). Road Traffic Safety Management Systems: Requirements with Guidance for Use. In: ISO TC 241/SC N9 Working Document Version 1, Stockholm.

International Traffic Safety Data and Analysis Group, (2012). IRTAD Road Safety Annual Report 2011, Organization for Economic Co-operation and Development/International Transport Forum, Paris. www.irtad.net.

Inuiguchi, M., Ichihasi, H. & Tanaka, H., (1990). Fuzzy Programming: A Survey of Recent Developments, In: Slowinski, R. & Teghem J. (Eds.), Stochastic Versus Fuzzy Approaches to Multi-objective Mathematical Programming under Uncertainty, Kluwer Academic Publishers, Dordrecht.

Jenkins L. & Anderson M., (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis, European Journal of Operational Research, Vol. 147, pp. 51-61.

Jiang, N. & Gruenwald, L., (2008). Estimating Missing Data in Data Streams. In: Kotagiri, R., Krishna, P.R., Mohania, M. & Nantajeewarawat E. (Eds.), Advances in Databases: Concepts, Systems and Applications, Vol. 4443, pp. 981–987.

Kabak, Ö. & Ruan, D., (2011). A cumulative belief degree-based approach for missing values in nuclear safeguards evaluation, IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 10, pp. 1441-1454.

Kahraman, C. & Tolga, E., (1998). Data envelopment analysis using fuzzy concept, In Proceedings of the 28th International Symposium on Multiple-Valued Logic, Los Alamitos, pp. 338-343.

Kanbur, R., (1987). Shadow Pricing, In: Eatwell, J., Milgate, M. & Newman, P. (Eds.), The New Palgrave: A Dictionary of Economics, Vol. 4, pp. 316-317.

Kao, C., (2008). A linear formulation of the two-level DEA model, Omega, International Journal of Management Science, Vol. 36, pp. 958-962.

Kao, C. & Liu, S.T., (2000). Fuzzy efficiency measures in data envelopment analysis, Fuzzy Sets and Systems, Vol. 113, pp. 427-437.

Keehley, P., Medlin, S., MacBride, S. & Longmire, L., (1997). Benchmarking for Best Practices in the Public Sector, Jossey Bass, California.

Koornstra, M., Lynam, D., Nilsson, G., Noordzij, P., Petterson, H-E., Wegman, F., Wouters, P., (2002). SUNflower: A Comparative Study of the Development of Road Safety in Sweden, the United Kingdom, and the Netherlands. SWOV Institute for Road Safety Research, Leidschendam.

Kumbhakar, S.C. & Lovell, C.A.K, (2000). Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

Kweon, Y. & Kockelman, K.M., (2005). The safety effects of speed limit changes: Use of panel models, including speed, use and design variables, 84th annual meeting of the Transportation Research Board, Washington.

Lai, M., Huang, H. & Wang, W., (2011). Designing a knowledge-based system for benchmarking: A DEA approach, Knowledge-Based Systems, Vol. 24, pp. 662-671.

Land Transport Safety Authority (LTSA), (2000). Road Safety Strategy 2010; A Consultation Document, National Road Safety Committee, LTSA, Wellington. http://www.ltsa.govt.nz/publications/rs-framework.html.

Lau, H.C.W., Lau, P.K.H., Fung, R.Y.K., Chan, F.T.S. & Ip, R.W.L., (2005). A virtual case benchmarking scheme for vendors' performance assessment, Benchmarking: An International Journal, Vol. 12, No. 1, pp. 61-80.

León, T., Liern, V., Ruiz, J.L. & Sirvent, I., (2003). A fuzzy mathematical programming approach to the assessment of efficiency with DEA models, Fuzzy Sets and Systems, Vol. 139, pp. 407-419.

Lertworasirikul, S., (2001). Fuzzy Data Envelopment Analysis for Supply Chain Modelling and Analysis. Dissertation Proposal in Industrial Engineering, North Carolina State University.

Lertworasirikul, S., Fang, S.C., Joines, J.A. & Nuttle, H., (2003). Fuzzy data envelopment analysis: A possibility approach, Fuzzy Sets and Systems, Vol. 139, pp. 379-394.

Lewin, A.Y., Morey, C.R. & Cook, T.J., (1982). Evaluating the administrative efficiency of courts, Omega, International Journal of Management Science, Vol. 10, No. 4, pp. 401-411.

Li, J., (2006). Robust rule-based prediction, IEEE Transactions on Knowledge and Data Engineering, Vol. 18, No. 8, pp. 1043-1054.

Lie, A. & Tingvall, C., (2000). How does EuroNCAP results correlate to real life injury risks: A paired comparison study of car-to-car crashes, In Proceedings of the IRCOBI conference, Montpellier.

Lindo Systems Inc., (2007). Optimization Modeling with LINGO, Lindo Systems Inc.

Little, R.J.A. & Rubin D.B., (1987). Statistical Analysis with Missing Data, John Wiley & Sons, New York.

Liu, H., Shah, S. & Jiang, W., (2004). On-line outlier detection and data cleaning, Computers and Chemical Engineering, Vol. 28, pp. 1635-1647.

Luu, V., Kim, S. & Huynh, T., (2008). Improving project management performance of large contractors using benchmarking approach, International Journal of Project Management, Vol. 26, No. 7, pp. 758-769.

Lynam, D., Hummel, T., Barker, J. & Lawson, S.D., (2004). European Road Assessment Programme EuroRAP I (2003) Technical Report. The AA Motoring Trust, Farnborough.

Malmquist, S., (1953). Index numbers and indifference surfaces, Trabajos de Estatistica, Vol. 4, pp. 209-242.

Meada, Y., Entani, T. & Tanaka, H., (1998). Fuzzy DEA with interval efficiency, In Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing, Aachen, Vol. 2, pp. 1067-1071.

Melyn W. & Moesen W., (1991). Towards a synthetic indicator of macroeconomic performance: Unequal weighting when limited information is available, Public Economic Research Paper 17, CES, KU Leuven.

Meng, W., Zhang, D., Qi, L. & Liu, W., (2008). Two-level DEA approaches in research evaluation, Omega, International Journal of Management Science, Vol. 36, pp. 950-957.

Mock, C.N., nii-Amon-Kotei, D. & Maier, R.V., (1997). Low utilization of formal medical services by injured persons in a developing nation: Health service data underestimate the importance of trauma, Journal of Trauma, Vol. 42, pp. 504-513.

Molnar, F., Hutton, B. & Fergusson, D., (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? Canadian Medical Association, Vol. 179, No. 8, pp. 751-753.

Morsink, P., Oppe, S., Reurings, M. & Wegman, F., (2007). Development of footprint methodology for road safety, Transportation Research Record: Journal of the Transportation Research Board, Vol. 2009, pp. 104-112.

Munda, G., (2005). Multi-Criteria Decision Analysis and Sustainable Development, In: Figueira, J., Greco, S. & Ehrgott, M. (Eds.), Multiple-Criteria Decision Analysis: State of the Art Surveys, Springer International

Series in Operations Research and Management Science, New York, pp. 953-986.

Noland, R.B., (2004). A Review of the Impact of Medical Care and Technology in Reducing Traffic Fatalities, IATSS Research, Vol. 28, No. 2, pp. 6-12.

Oltman, L.B. & Yahia, S.B., (2008). Yet Another Approach for Completing Missing Values, In: Yahia, S.B., Nguifo, E.M. & Belohlavek, R. (Eds.), Concept Lattices and Their Applications, Springer-Verlag, Germany, pp. 155-169.

Organization for Economic Co-operation and Development (OECD), (2002). Road Safety: What's the Vision? OECD, Paris.

Organization for Economic Co-operation and Development (OECD), (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide, OECD, Paris. www.oecd.org/publishing/corrigenda.

Organization for Economic Co-operation and Development/International Transport Forum (OECD/ITF), (2006). Country Reports on Road Safety Performance, Joint Transport Research Centre of the OECD and the ITF, Paris.

Organization for Economic Co-operation and Development/International Transport Forum (OECD/ITF), (2008a). Towards Zero: Ambitious Road Safety Targets and the Safe System Approach, Joint Transport Research Centre of the OECD and the ITF, Paris.

Organization for Economic Co-operation and Development/International Transport Forum (OECD/ITF), (2008b). Country Reports on Road Safety Performance, Joint Transport Research Centre of the OECD and the ITF, Paris.

Organization for Economic Co-operation and Development/International Transport Forum (OECD/ITF), (2011a). Transport Outlook 2011: Meeting the Needs of 9 Billion People, Joint Transport Research Centre of the OECD and the ITF, Paris.

Organization for Economic Co-operation and Development/International Transport Forum (OECD/ITF), (2011b). Reporting on Serious Road Traffic Casualties: Combining and Using Different Data Sources to Improve

Understanding of Non-fatal Road Traffic Crashes. Joint Transport Research Centre of the OECD and the ITF, Paris.

O'Reilly, D., Hopkin, J., Loomes, G., Jones-Lee, M., Philips, P., McMahon, K., Ives, D., Soby, B., Ball, D. & Kemp, R., (1994). The value of road safety: UK research on the valuation of preventing non-fatal injuries, Journal of Transport Economics and Policy, Vol. 28, No. 1, pp. 45-59.

Pastor, J.T. & Lovell, C.A.K., (2007). Circularity of the Malmquist productivity index, Economic Theory, Vol. 33, pp. 591-599.

Peat, J. & Bartion, B., (2005). Medical Statistics: A Guide to Data Analysis and Critical Appraisal, Blackwell Publishing Ltd., Oxford.

Pospiech-Kurkowska, S., (2008). Processing of missing data in a fuzzy system. In: Pietka E. & Kawa J. (Eds.), Information Technology in Biomedicine, Springer, Germany, pp. 453-460.

Ramanathan, R., (2006). Evaluating the comparative performance of countries of the Middle East and North Africa: A DEA application, Socio-Economic Planning Sciences, Vol. 40, pp. 156-167.

Rodríguez, R., Martínez, L., Ruan, D. & Liu, J., (2010). Using collaborative filtering for dealing with missing values in nuclear safeguards evaluation, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 18, No. 4, pp.431-449.

Rousseeuw, P. & Leory, A., (1987). Robust Regression and Outlier Detection, Wiley Series in Probability and Statistics.

Rubin, D.B., (1976). Inference and missing data, Biometrika, Vol. 63, pp. 581-592.

Rubin, D.B., (1987). Multiple Imputation for Non-response in Surveys, John Wiley & Sons, New York.

Rubin, D.B., (1996). Multiple imputation after 18+ years, Journal of the American Statistical Association, Vol. 91, pp. 473-489.

Rumar, K., (1982). The human factor in road safety, Invited paper at 11th Australian Road Research Board Conference, Melbourne.

Rumar, K., (1985). The Role of Perceptual and Cognitive Filters in Observed Behavior, In: Evans, L. & Schwing, R. (Eds.), Human Behavior in Traffic Safety, Plenum Press.

Saati, S. & Memariani, A., (2005). Reducing weight flexibility in fuzzy DEA, Applied Mathematics and Computation, Vol. 161, No. 2, pp. 611–622.

Saati, S., Memariani, A. & Jahanshahloo, G.R., (2002). Efficiency analysis and ranking of DMUs with fuzzy data, Fuzzy Optimization and Decision Making, Vol. 1, pp. 255-267.

Saaty, T.L., (1980). The Analytic Hierarchy Process, McGraw-Hill, New York.

Saisana, M., & Tarantola, S., (2002). State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development, EUR 20408 EN Report, the Joint Research Center of European Commission, Ispra.

Saisana, M., Tarantola, S. & Saltelli, A., (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators, Journal of the Royal Statistical Society A, Vol. 168, No. 2, pp. 307-323.

Sarrico, C.S. & Dyson, R.G., (2004). Restricting virtual weights in data envelopment analysis, European Journal of Operational Research, Vol. 159, pp. 17-34.

SARTRE 3 consortium, (2004). European Drivers and Road Risk, Inrets.

Schafer, J.L. & Graham, J.W., (2002). Missing data: Our view of the state of the art, Psychological Methods, Vol. 7, No. 2, pp. 147-177.

Schafer, J.L. & Olsen M.K., (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective, Multivariate Behavioral Research. Vol. 33, pp. 545-571.

Schefcyzk, M., (1993). Industrial benchmarking: A case study of performance analysis techniques. International Journal of Production Economics, Vol. 32, No. 1, pp. 1-11.

Schiffler, R.E., (1988). Maximum Z score and outliers, The American Statistician, Vol. 42, No.1, pp. 79-80.

Seiford, L.M. & Thrall, R.M., (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. Journal of Econometrics, Vol. 46, pp. 7-38.

Sengupta, J.K., (1992). A fuzzy systems approach in data envelopment analysis, Computers and Mathematics with Applications, Vol. 24, No. 8, pp. 259-266.

Sexton, T.R., Silkman, R.H. & Hogan, A.J., (1986). Data Envelopment Analysis: Critique and Extensions. In: Silkman, R.H. (Ed.), Measuring Efficiency: An Assessment of Data Envelopment Analysis, Jossey-Bass, San Francisco, pp. 73-104.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K., (2011). A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: A case study in road safety performance evaluation, Expert systems with applications, Vol. 38, No. 12, pp. 15262-15272.

Shen, Y., Li, T., Hermans, E., Ruan, D., Wets, G., Vanhoof, K. & Brijs, T., (2010). A hybrid system of neural networks and rough sets for road safety performance indicators, Soft Computing, Vol. 14, No. 12, pp. 1255-1263.

Silva-Ramírez, E.-L., Pino-Mejías, R., López-Coello, M. & Cubiles-de-la-Vega, M.-D., (2011). Missing value imputation on missing completely at random data using multilayer perceptrons, Neural Networks, Vol. 24, pp. 121-129.

Simar, L. & Wilson, P.W., (2000). Statistical inference in non-parametric frontier models: The state of the art, Journal of Productivity Analysis, Vol. 13, pp. 49-78.

Singh, R.K., Murty, H.R., Gupta, S.K. & Dikshit, A.K., (2009). An overview of sustainability assessment methodologies, Ecological Indicators, Vol. 9, pp. 189-212.

Smirlis, Y.G., Maragos, E.K. & Despotis, D.K., (2006). Data envelopment analysis with missing values: An interval DEA approach, Applied Mathematics and Computation, Vol. 177, pp. 1-10.

Spendolini, M., (1992). The Benchmarking Book, American Management Association, New York.

SPSS Inc., (2007). SPSS Missing Values™ 17.0. http://www.spss.com.

Srinivas, T. (2000). A benchmarking method for business-process reengineering and improvement, The International Journal of Flexible Manufacturing Systems, Vol. 12, pp. 291–304.

Taylor, M.C., Lynam, D.A. & Baruya, A., (2000). The effect of drivers' speed on the frequency of road accidents, TRL Report 421, Transport Research Laboratory, Crowthorne.

Thanassoulis, E., Portela, M.C. & Allen, R., (2004). Incorporating Value Judgements in DEA, In: Cooper, W.W., Seiford, L.M. & Zhu, J. (Eds.), Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston, pp. 99-138.

Thomas, P., Yannis, G., Duchamp, G., Vis, M., Jahi, H., Vallet, G. & Dupont, E., (2009). SafetyNet Final Activity Report, the EU FP6 project SafetyNet.

Thompson, R.G., Langemeier, L., Lee, C., Lee, E. & Thrall, R., (1990). The role of multiplier bounds in efficiency analysis with application to Kansas farming, Journal of Econometrics, Vol. 46, pp. 93-108.

Tlig, H. & Rebai, A., (2009). A mathematical approach to solve data envelopment analysis models when data are LR fuzzy numbers, Applied Mathematical Sciences, Vol. 3, No. 48, pp. 2383-2396.

Tortosa-Ausina, E., Grifell-Tatje, E., Armero, C. & Conesa, D., (2008). Sensitivity analysis of efficiency and Malmquist productivity indices: An application to Spanish savings banks, European Journal of Operational Research, Vol. 184, pp. 1062-1084.

Transportation Research Board (TRB), (1998). Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits, Special report 254, TRB, National Academy Press, Washington DC.

Traynor, T.L., (2008). Regional economic conditions and crash fatality rates: A cross-county analysis, Journal of Safety Research, Vol. 39, pp. 33-39.

Traynor, T.L., (2009). The impact of state level behavioral regulations on traffic fatality rates, Journal of Safety Research, Vol. 40, No. 6, pp. 421-426.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stanisfer, R.L. & Castellan, N.J., (1977). Tri-level Study of the Causes of Traffic Accidents, Report No. DOT-HS-034-3-535-77 (TAC).

United Nations Economic Commission for Europe (UNECE), (2011). http://www.unece.org/.

Vercellis, C., (2009). Business Intelligence: Data Mining and Optimization for Decision Making, John Wiley & Sons, Ltd.

Vis, M.A. (Ed.), (2005). State of the Art Report on Road Safety Performance Indicators, Deliverable D3.1 of the EU FP6 project SafetyNet.

Vis, M.A. & Eksler, V. (Eds.), (2008). Road Safety Performance Indicators: Updated Country Comparisons, Deliverable D3.11a of the EU FP6 project SafetyNet.

Wang, Y.M., Greatbanks, R. & Yang, J.B., (2005). Interval efficiency assessment using data envelopment analysis, Fuzzy Sets and Systems, Vol. 153, No. 3, pp. 347-370.

Wang H. & Wang, S., (2009). Discovering patterns of missing data in survey databases: An application of rough sets, Expert Systems with Applications, Vol. 36, pp. 6256-6260.

Wegman, F., Eksler, V., Hayes, S., Lynam, D., Morsink, P. & Oppe, S., (2005). SUNflower+6: A Comparative Study of the Development of Road Safety in the SUNflower+6 Countries: Final Report. SWOV Institute for Road Safety Research, Leidschendam.

Wegman, F. & Aarts L. (Eds.), (2006). Advancing Sustainable Safety: National Road Safety Outlook for 2005-2020, SWOV Institute for Road Safety Research, Leidschendam.

Wegman, F., Commandeur, J., Doveh, E., Eksler, V., Gitelman, V., Hakkert, S., Lynam, D. & Oppe, S., (2008). SUNflowerNext: Towards a Composite Road Safety Performance Index, Deliverable D6.16 of the EU FP6 project SafetyNet.

Wegman, F. & Oppe, S., (2010). Benchmarking road safety performances of countries, Safety Science, Vol. 48, pp. 1203–1211.

Williams, G.J., Baxter, R.A., He, H.X., Hawkins, S. & Gu, L., (2002). A comparative study of RNN for outlier detection in data mining, IEEE International Conference on Data-mining (ICDM'02), Maebashi, CSIRO Technical Report CMIS-02/102.

Wong, S.C., Sze, N.N., Yip, H.F., Loo, B.P.Y., Hung, W.T. & Lo, H.K., (2006). Association between setting quantified road safety targets and road fatality reduction, Accident Analysis and Prevention, Vol. 38, pp. 997-1005.

Wong, Y.H.B. & Beasley, J.E., (1990). Restricting Weight Flexibility in Data Envelopment Analysis, Journal of the Operational Research Society, Vol. 41, No. 9, pp. 829-835.

World Road Association (PIARC), (2003). Road Safety Manual, version 1.0, PIARC.

World Health Organization (WHO), (2004). World Report on Road Traffic Injury Prevention, WHO, Geneva. http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/.

World Health Organization (WHO), (2008). World Health Statistics 2008, WHO, Geneva. http://www.who.int/whosis/whostat/en/.

World Health Organization (WHO), (2009). Global Status Report on Road Safety - Time for Action, WHO, Geneva. http://www.who.int/violence_injury_prevention/road_safety_status/2009/en/.

Yang, J.B., Wong, B.Y.H., Xu, D.L. & Stewart, T.J., (2009). Integrating DEA-oriented performance assessment and target setting using interactive MOLP methods, European Journal of Operational Research, Vol. 195, pp. 205-222.

Yannis, G., Lejeune, P., Treny, V., Hemdorff, S., Bergel, R., Haddak, M., Holló, P., Cardoso, J., Bijleveld, F., Houwing, S. & Bjørnskau, T., (2008). Risk Exposure Data: Recommendations for Collection and Exploitation, Deliverable D2.5 of the EU FP6 project SafetyNet.

Yörük, B.K. & Zaim, O., (2005). Productivity growth in OECD countries: A comparison with Malmquist indices, Journal of Comparative Economics, Vol. 33, pp. 401-420.

Zadeh, L.A., (1965). Fuzzy sets, Information and Control, Vol. 8, pp. 338-353.

Zador, P.L., (1991). Alcohol-related relative risk of fatal driver injuries in relation to driver age and sex, Journal of Studies on Alcohol, Vol. 52, pp. 302-310.

Zhou, P., Ang, B.W. & Poh, K.L., (2007). A mathematical programming approach to constructing composite indicators, Ecological Economics, Vol. 62, pp. 291-297.

Zhu, J., (2003). Quantitative Models for Performance Evaluation and Benchmarking, Kluwer Academic Publishers, Boston.

Zhu, J. & Cook, W.D. (Eds.), (2007). Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis, Springer, New York.

Appendixes

Appendix I. The dendrograms of hierarchical clustering analysis

The dendrograms from three clustering techniques, i.e., the Ward's method, the Centroid Linkage method, and the Average Linkage (between and within groups) method in SPSS 17.0, are shown as follows:

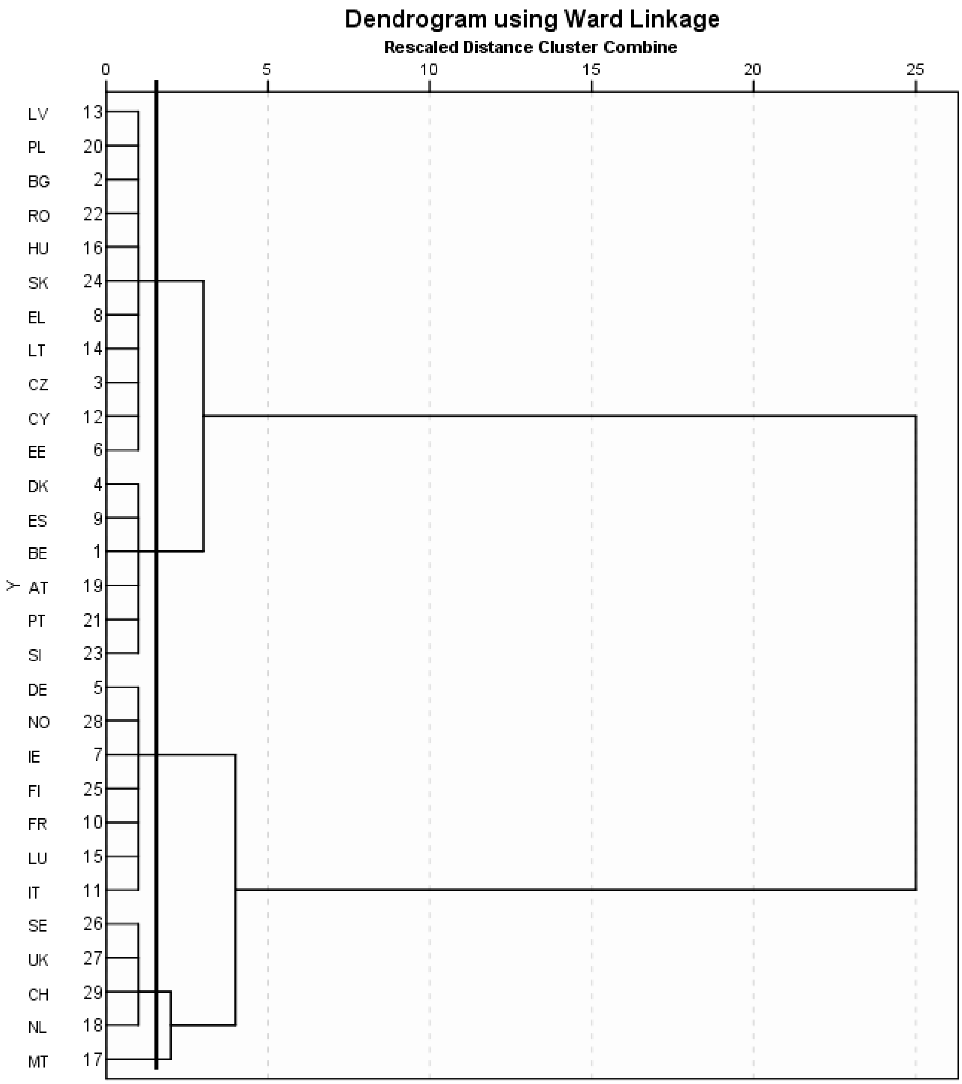


Figure AI-1 Dendrogram using the Ward's method

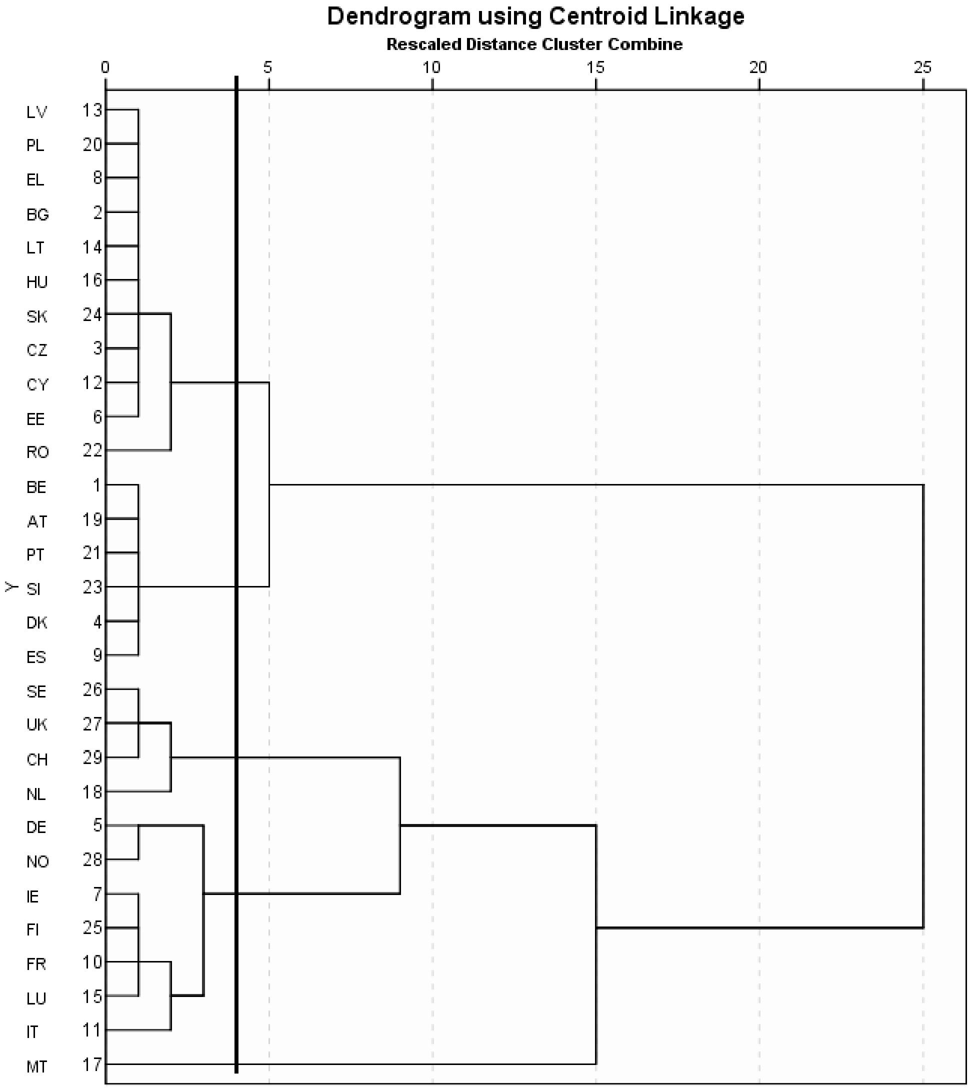


Figure AI-2 Dendrogram using the Centroid Linkage method

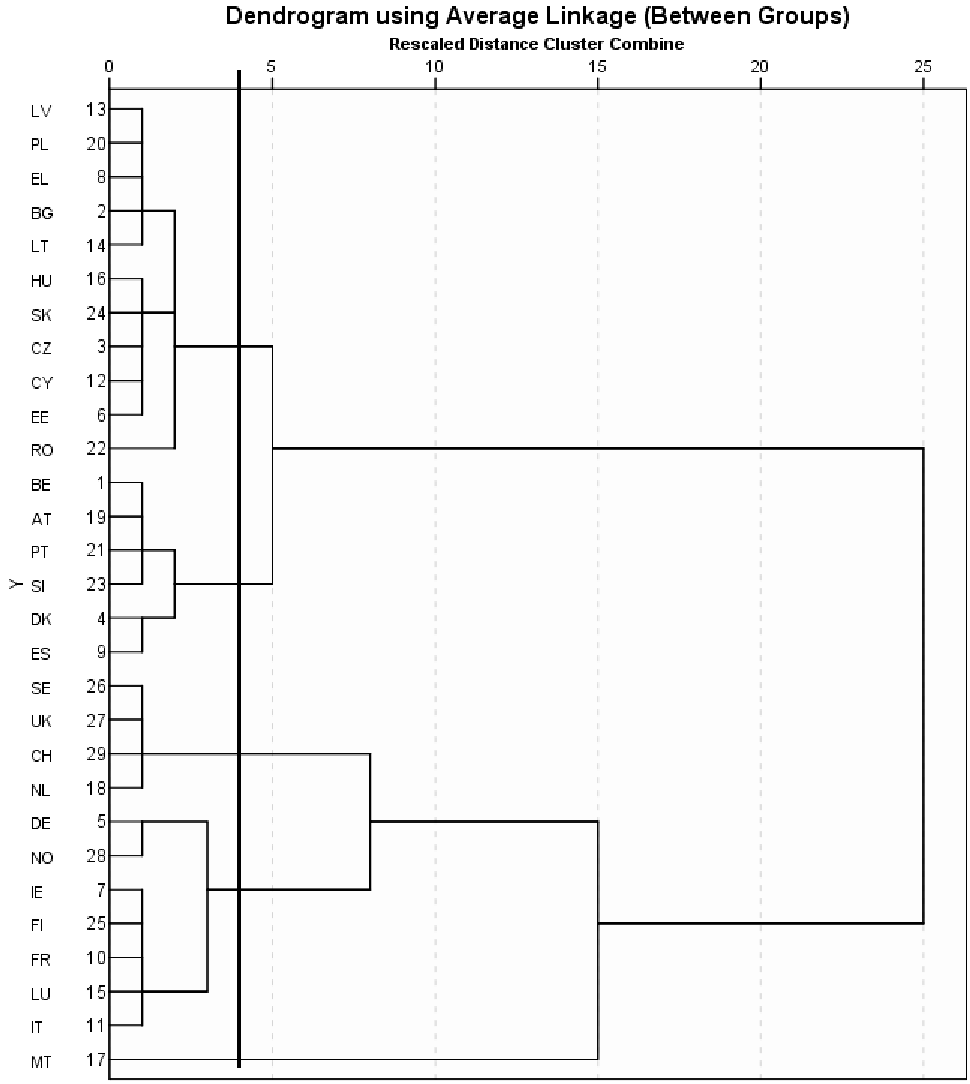


Figure AI-3 Dendrogram using the Average Linkage (between groups) method

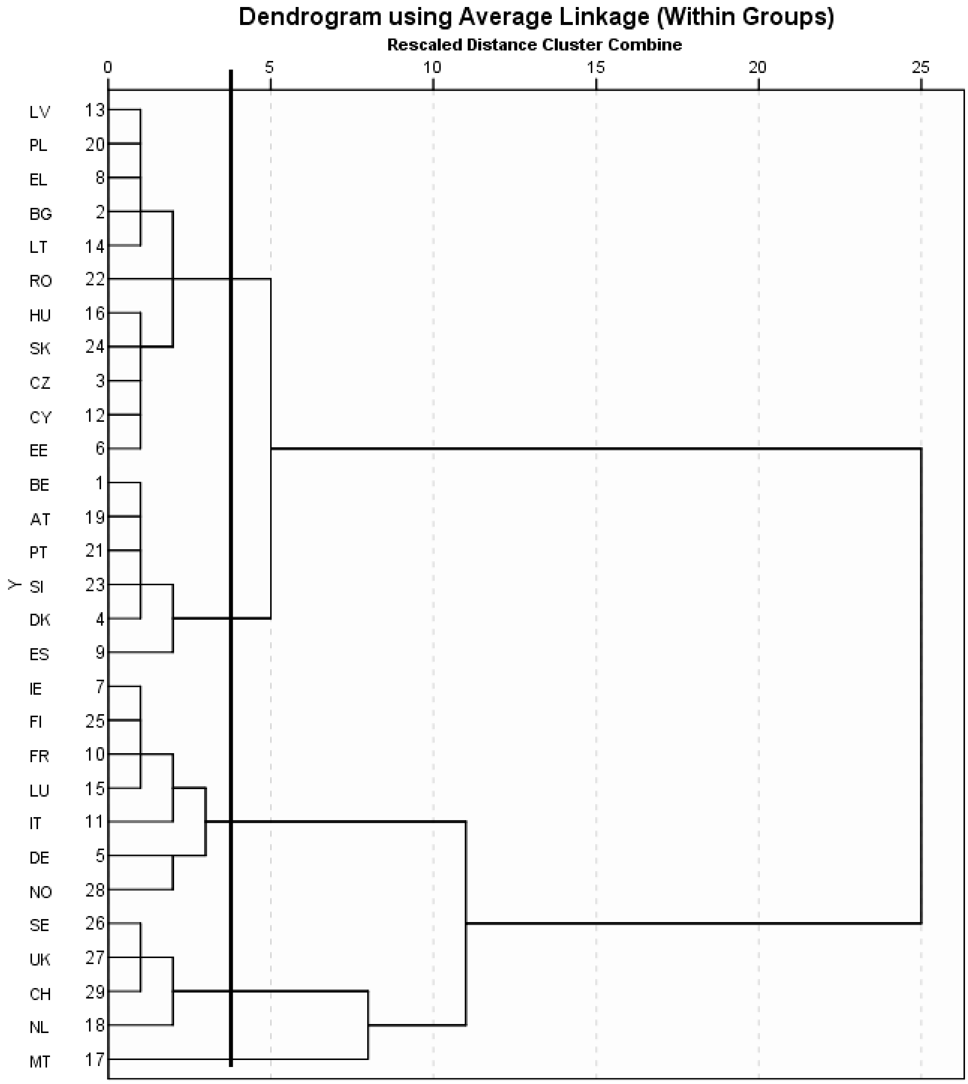
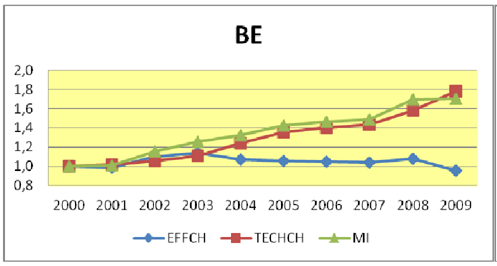
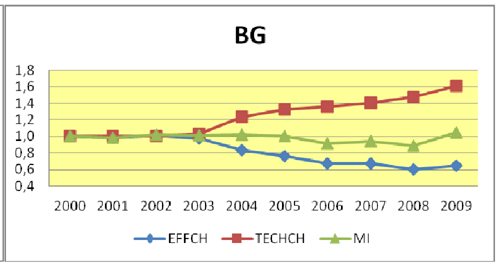


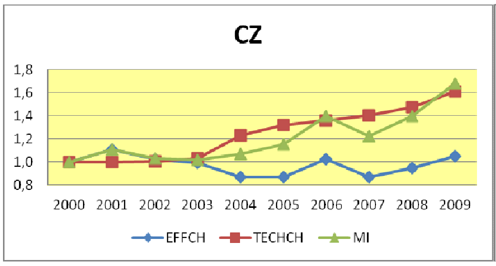
Figure AI-4 Dendrogram using the Average Linkage (within groups) method

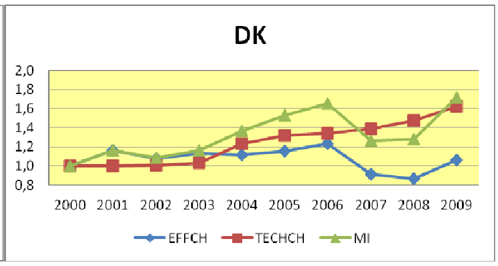
Appendix II. The evolution in *EFFCH*, *TECHCH*, and *MI* of each of the 28 European countries, 2000-2009

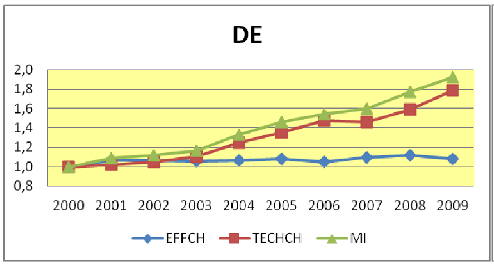
Based on the DEA-RS-MI model, the evolution in *MI* of each of the 28 European countries and its decomposition into technical and efficiency changes in 2000-2009 are shown as follows:

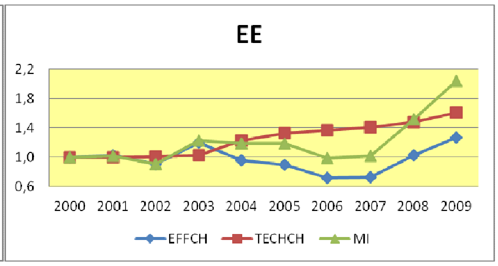


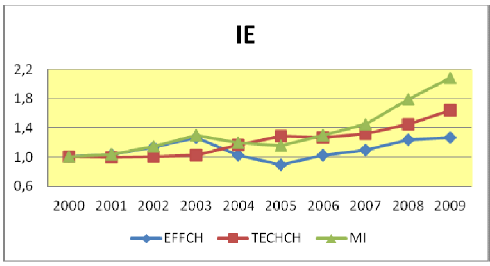


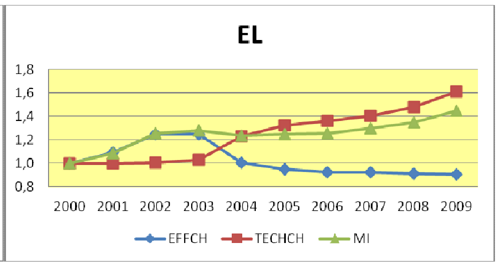


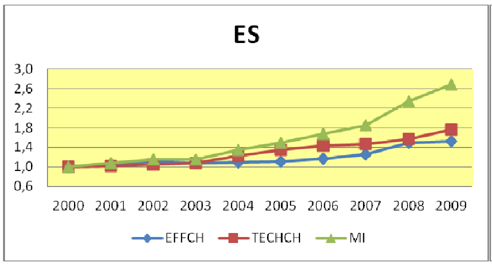
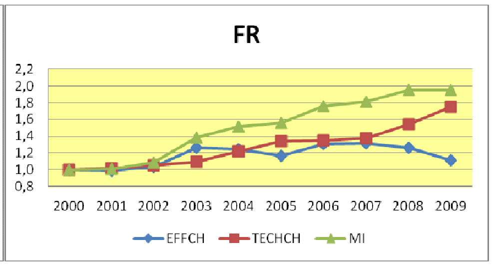
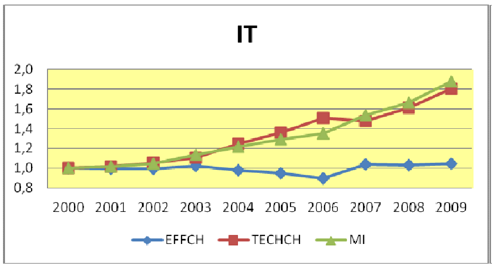
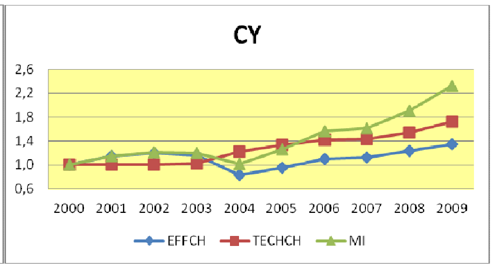
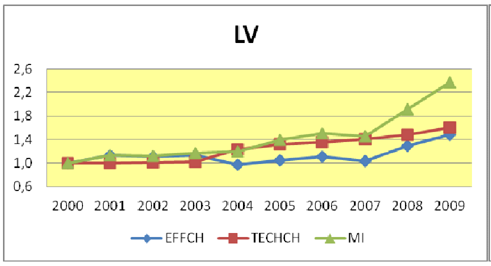
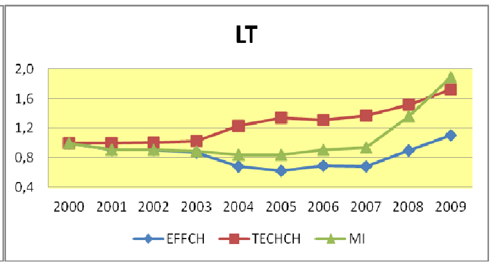
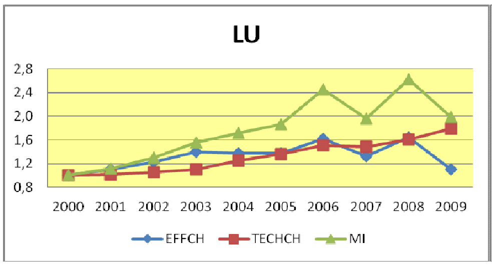
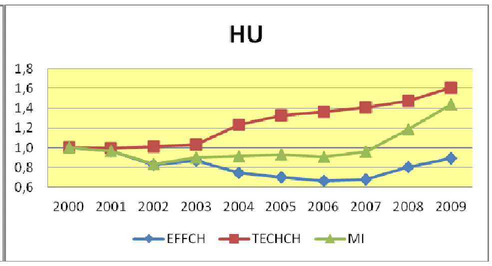
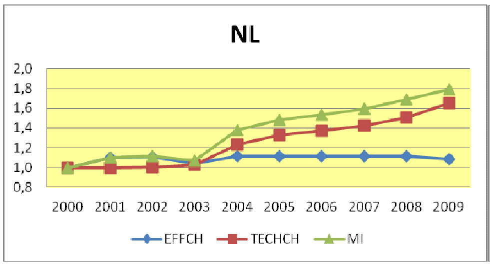
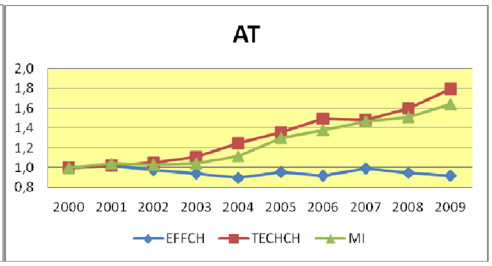


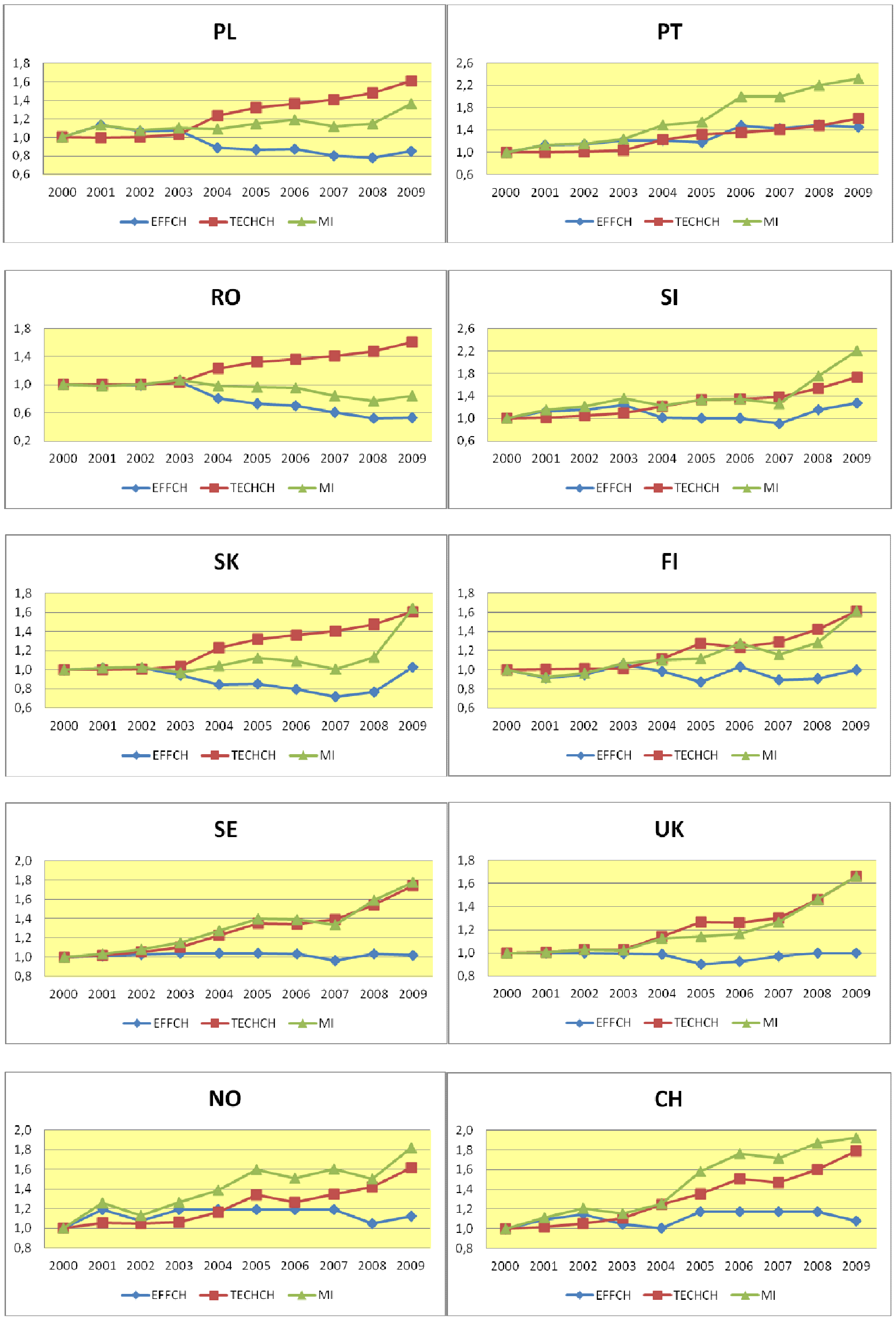




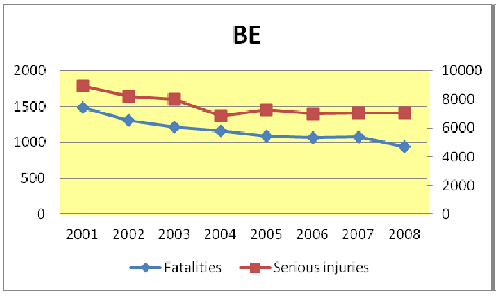


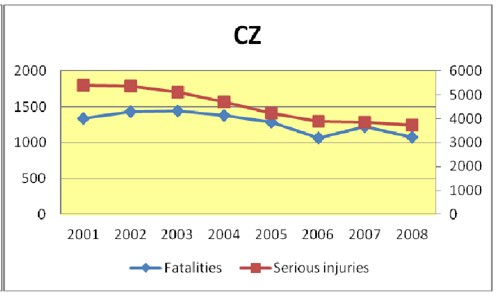


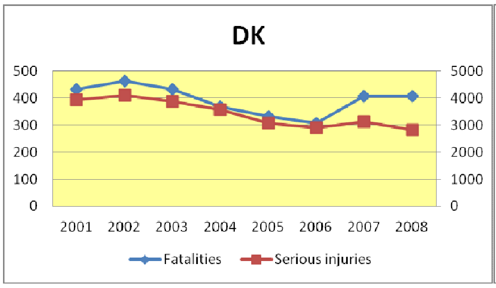

ES
3,0
2,6
2,2
1,8
1,4
1,0
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

FR
2,2
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

IT
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

CY
2,6
2,2
1,8
1,4
1,0
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

LV
2,6
2,2
1,8
1,4
1,0
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

LT
2,0
1,6
1,2
0,8
0,4
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

LU
2,8
2,4
2,0
1,6
1,2
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

HU
1,8
1,6
1,4
1,2
1,0
0,8
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

NL
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

AT
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

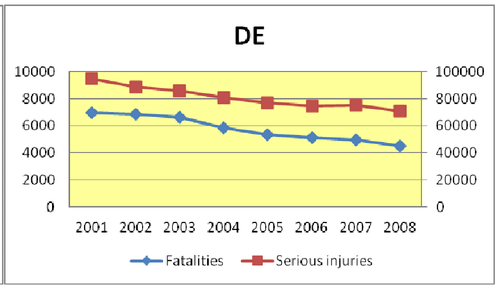

PL
1,8
1,6
1,4
1,2
1,0
0,8
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
PT
2,6
2,2
1,8
1,4
1,0
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
RO
1,8
1,4
1,0
0,6
0,2
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
SI
2,6
2,2
1,8
1,4
1,0
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
SK
1,8
1,6
1,4
1,2
1,0
0,8
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
FI
1,8
1,6
1,4
1,2
1,0
0,8
0,6
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
SE
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
UK
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
NO
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI
CH
2,0
1,8
1,6
1,4
1,2
1,0
0,8
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
EFFCH TECHCH MI

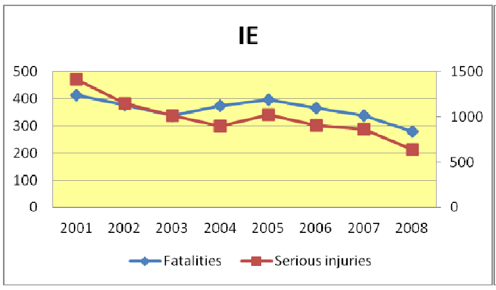
Appendix III. The evolution of the number of serious injuries and fatalities of the 15 European countries, 2001-2008

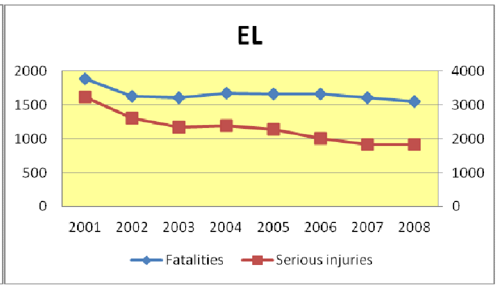


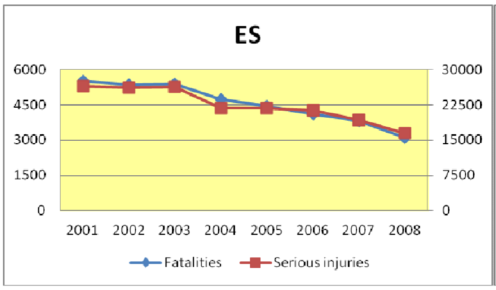


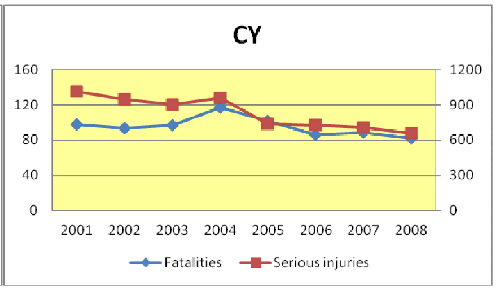


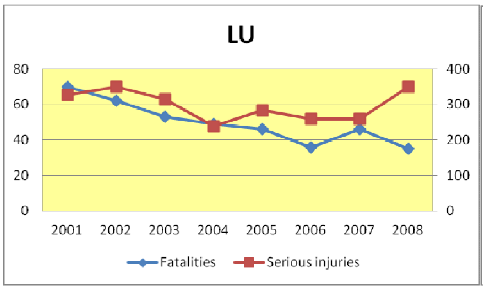


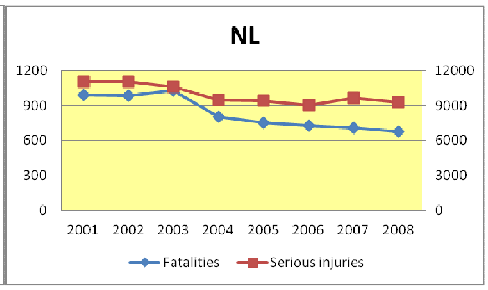


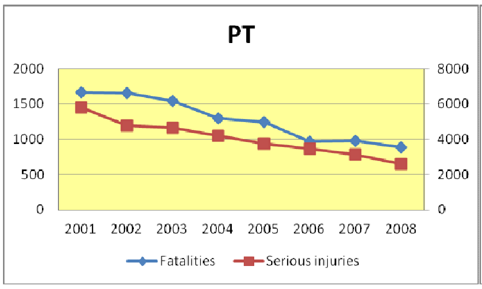


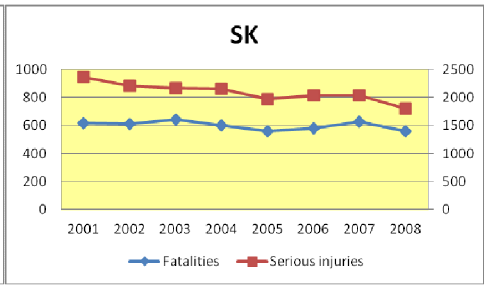


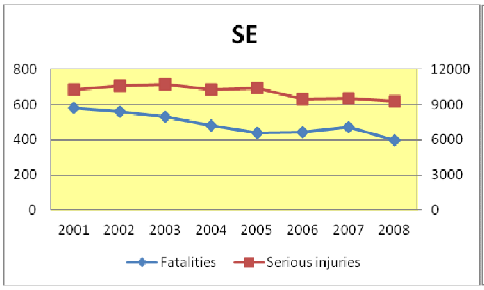


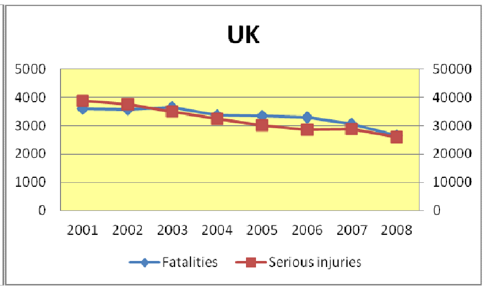

LU
80
60
40
20
0
400
300
200
100
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries

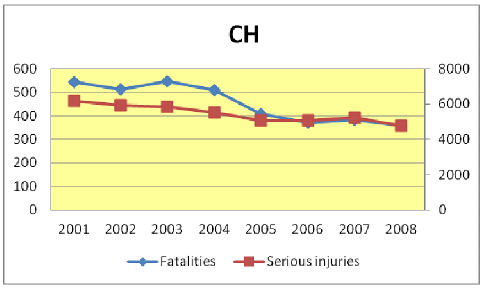

NL
1200
900
600
300
0
12000
9000
6000
3000
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries


PT
2000
1500
1000
500
0
8000
6000
4000
2000
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries


SK
1000
800
600
400
200
0
2500
2000
1500
1000
500
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries


SE
800
600
400
200
0
12000
9000
6000
3000
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries


UK
5000
4000
3000
2000
1000
0
50000
40000
30000
20000
10000
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries


CH
600
500
400
300
200
100
0
8000
6000
4000
2000
0
2001 2002 2003 2004 2005 2006 2007 2008
Fatalities
Serious injuries

Memberships and Publications (2008-2012)

MEMBERSHIPS

- Program Committee Member of the 2011 International Conference on Intelligent Systems and Knowledge Engineering (ISKE'11) http://iske2011.sjtu.edu.cn/Committees.html
- International Scientific Committee Membership of the 10th International FLINS Conference on Uncertainty Modeling in Knowledge Engineering and Decision Making (FLINS'12) http://www.flins2012.itu.edu.tr/scientific_committee.html

JOURNAL ARTICLES

[1] **Shen, Y.**, Li, T., Hermans, E., Ruan, D., Wets, G., Vanhoof, K. & Brijs, T. (2010). A hybrid system of neural networks and rough sets for road safety performance indicators. *Soft Computing*, Vol. 14, No. 12, pp. 1255-1263.

[2] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K. (2010). Evaluating trauma management performance in Europe: A Multiple-Layer Data Envelopment Analysis Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2, No. 2148, pp. 69-75.

[3] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K. (2011). A generalized multiple layer data envelopment analysis model for hierarchical structure assessment: A case study in road safety performance evaluation. *Expert systems with applications*, Vol. 38, No. 12, pp. 15262-15272.

[4] **Shen, Y.**, Ruan, D., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K. (2011). Sustainable road transport in the European Union: Changes in undesirable impacts. *Transportation Research Record: Journal of the*

Transportation Research Board, No. 2242, pp. 37-44.

[5] **Shen, Y.**, Ruan, D., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K. (2011). Modeling qualitative data in data envelopment analysis for composite indicators. *International Journal of Systems Assurance Engineering and Management*, Vol. 2, No. 1, pp. 21-30.

[6] Bao, Q., Ruan, D., **Shen, Y.**, Hermans, E. & Janssens, D. (2012). Improved hierarchical fuzzy TOPSIS for road safety performance evaluation. *Knowledge-Based Systems*. doi: 10.1016/j.knosys.2011.08.014.

[7] **Shen, Y.**, Hermans, E., Brijs, T., Wets, G. & Vanhoof, K. (2012). Road safety risk evaluation and target setting using data envelopment analysis and its extensions. *Accident Analysis & Prevention,* doi: 10.1016/j.aap.2012.02.020.

[8] Li, T., Ruan, D., **Shen, Y.**, Hermans, E. & Wets, G.. A new weighing approach based on rough sets theory and granular computing for road safety indicator analysis. *Computational Intelligence* (*1st revision*).

CONFERENCE PROCEEDINGS

[9] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Vanhoof, K. & Brijs, T. (2008). Development of a composite road safety performance indicator based on neural networks. In *Proceedings of the 3rd International Conference on Intelligent Systems and Knowledge Engineering (ISKE'08)*, Xiamen, IEEE press, Vol. 2, pp. 901-906.

[10] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K. (2009). Road safety performance evaluation based on a multiple layer data envelopment analysis model. In *Proceedings of the 4th International IRTAD Conference on Road Safety*, Seoul, pp. 315-324.

[11] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Vanhoof, K. & Brijs, T. (2009). A generalized multiple layer data envelopment analysis model. In *Proceedings of the 4th International Conference on Intelligent Systems and Knowledge Engineering (ISKE'09)*, Diepenbeek, World Scientific Proceedings Series on Computer Engineering and Information Science,

Vol. 2, pp. 215-220.

[12] Li, T., **Shen, Y.**, Ruan, D., Hermans, E. & Wets, G. (2009). Integrating rough sets with neural networks for weighting road safety performance indicators. In *Proceedings of the 4th International Conference on Rough Set and Knowledge Technology (RSKT'09)*, Gold Coast, Lecture Notes in Artificial Intelligence, Vol. 5589, pp. 60-67.

[13] **Shen, Y.**, Hermans, E., Ruan, D., Wets, G., Brijs, T. & Vanhoof, K. (2010). Evaluating trauma management performance in Europe. In *Proceedings of The 89th Annual Meeting of the Transportation Research Board (TRB'10) CD-ROM*, Washington D.C..

[14] **Shen, Y.**, Hermans, E., Brijs, T., Ruan, D., Wets, G. & Vanhoof, K. (2010). Road safety evaluation and target setting using data envelopment analysis. In *Proceedings of the 15th International Conference Road Safety on Four Continents (RS4C'10) CD-ROM*, Abu Dhabi, pp. 282-293.

[15] **Shen, Y.**, Hermans, E., Ruan, D., Vanhoof, K., Brijs, T. & Wets, G. (2010). A DEA-based Malmquist productivity index approach in assessing road safety performance. In *Proceedings of the 9th International FLINS Conference*, Chengdu, World Scientific Proceedings Series on Computer Engineering and Information Science, Vol. 4, pp. 923-928.

[16] Bao, Q., Ruan, D., **Shen Y.** & Hermans, E. (2010). Creating a composite road safety performance index by a hierarchical fuzzy TOPSIS approach. In *Proceedings of the 5th International Conference on Intelligent Systems and Knowledge Engineering (ISKE'10)*, Hangzhou, IEEE press, pp. 458-463.

[17] **Shen, Y.**, Ruan, D., Hermans, E., Brijs, T., Wets, G. & Vanhoof, K. (2011). Changes in undesirable impacts on sustainable road transport of a set of European countries. In *Proceedings of the 90th Annual Meeting of the Transportation Research Board (TRB'11) CD-ROM*, Washington D.C..

[18] **Shen, Y.**, Hermans, E., Ruan, D., Brijs, T., Wets, G. & Vanhoof, K. (2011). Data envelopment analysis in road safety: Model building

considerations. Abstract in *Proceedings of the 12th congress of the French National Society of Operations Research and Decision Science (ROADEF'11) USB-Disk*, Saint-Étienne.

[19] **Shen, Y.**, Hermans, E., Ruan, D., Brijs, T., Wets, G. & Vanhoof, K. (2011). Development of a multiple layer DEA model with applications in road safety performance evaluation. Abstract in *Proceedings of the 9th International Conference on Data Envelopment Analysis (DEA'11) USB-Disk*, Thessaloniki.

[20] Hermans, E., **Shen, Y.**, Brijs, T., Wets, G. & Vanhoof, K. (2011). Road safety target setting in Europe based on a combined data envelopment analysis and time series analysis approach. Abstract in *Proceedings of the 9th International Conference on Data Envelopment Analysis (DEA'11) USB-Disk*, Thessaloniki.

BOOK CHAPTERS

[21] **Shen, Y.**, Hermans, E., Brijs, T., Wets, G. & Vanhoof, K.. Investigating the progress towards sustainable road transport in Europe: Lessons learned from a DEA-based Malmquist productivity index. (*submitted*).

[22] Bao, Q., Ruan, D., **Shen Y.**, Hermans, E. & Janssens, D. (2012). Fuzzy multi-criteria decision making: Applications for road safety performance evaluation. In: Kahraman, C. (Ed.), *Computational Intelligence Applications in Industrial Engineering*, Atlantis Springer. (*Accepted*).

REPORTS

[23] Wilmots B., **Shen Y.**, Hermans E. & Ruan D. (2011). Missing data treatment – Overview of possible solutions. Diepenbeek: Policy Research Centre Mobility and Public Works, track Traffic Safety, RA-MOW-2011-002, pp. 1-35.

Samenvatting

Verkeersslachtoffers en verkeersdoden worden tegenwoordig beschouwd als één van de belangrijkste volksgezondheidsaangelegenheden waarvoor inspanningen met het oog op een efficiënte en duurzame preventie vereist zijn. Omdat steeds meer landen maatregelen nemen om hun verkeersveiligheidssituatie te verbeteren, is er een groeiende behoefte voor landen om hun eigen verkeersveiligheidsprestaties te evalueren, om deze te vergelijken met die van andere landen, en om bovendien te leren van “goede landen” bij de ontwikkeling van hun eigen verkeersveiligheidsbeleid. Dit doctoraatsonderzoek richt zich op het benchmarken van het verkeersveiligheidsproduct enerzijds en het verkeersveiligheidsprogramma anderzijds op basis van indicatoren gerelateerd aan het verkeersveiligheidsrisico en de verkeersveiligheidsprestatie voor 28 Europese landen. De data envelopment analyse (DEA) techniek, die oorspronkelijk ontwikkeld werd om de relatieve efficiëntie te beoordelen van een homogene set van eenheden op basis van meerdere inputs en outputs, werd bestudeerd en toegepast doorheen dit proefschrift. Meerdere uitbreidingen van de methodologie werden onderzocht en voorgesteld om te beantwoorden aan de specifieke onderzoeksvragen. Deze benchmark studie op vlak van verkeersveiligheid verschafte ons nuttige inzichten waardoor waardevolle aanbevelingen met betrekking tot verkeersveiligheid aan beleidsmakers gegeven konden worden, bijvoorbeeld door te wijzen op haalbare doelstellingen en het formuleren van aandachtspunten om het verkeersveiligheidsniveau te verbeteren.

Bij het benchmarken van het verkeersveiligheidsproduct lag de nadruk op verscheidene finale verkeersveiligheidsuitkomsten (zoals dodelijke slachtoffers). Hierbij werden landen vergeleken wat betreft hun verkeersveiligheidsrisico gedefinieerd op basis van verschillende blootstellingsmaten, evenals de evolutie hierin over de tijd. Meer specifiek ontwikkelden we een DEA-gebaseerd verkeersveiligheidsmodel (DEA-RS) om de globale verkeersveiligheid van de 28 Europese landen te evalueren door tezelfdertijd drie belangrijke risico-indicatoren in rekening te brengen (zijnde het aantal doden per miljoen inwoners, het aantal doden per 10 miljard afgelegde personenkilometers, en het aantal doden per miljoen personenwagens). Op deze manier konden we de

‘efficiëntie’ van elk land identificeren. Na een clusteranalyse om landen die inherente gelijkenissen vertonen te groeperen, pasten we verder een categorisch DEA-RS model toe om de best presterende en ondermaats presterende landen in elke groep te identificeren. Zo konden we nuttige voorbeeldlanden identificeren, alsook een reeks praktische doelstellingen met betrekking tot verkeersdoden bepalen voor de landen die ondermaats presteren.

Om bovendien de dynamische verkeersveiligheidsontwikkeling in elk land te vatten, pasten we de Malmquist productiviteitsindex toe om veranderingen in de verkeersveiligheidsprestaties van landen doorheen de tijd te beoordelen. Hierbij keken we niet enkel naar de evolutie van de finale uitkomsten op vlak van verkeersveiligheid binnen een bepaalde periode, maar ook naar de veranderingen in blootstelling tijdens dezelfde periode. Bijgevolg leverde dit objectievere resultaten op dan de resultaten die gebaseerd zijn op de traditionele indicator, waar enkel procentuele veranderingen in het aantal verkeersdoden gemeten worden. De opdeling van de index in efficiëntie verandering (of “catch-up” effect) en technische verandering (of “frontier-shift” effect) verstrekte bovendien waardevolle informatie over het feit of de verkeersveiligheidsverbetering in elk land werd bereikt door een land-specifieke vooruitgang ten opzichte van de andere landen die werden beoordeeld of enkel door een algemene verbetering op technologisch vlak.

Bovendien onderzochten we in het kader van verkeersveiligheidsproduct benchmarking de mogelijkheid om ook het aantal zwaargewonden op te nemen als extra indicator van de finale uitkomsten van verkeersveiligheid en analyseerden we de impact hiervan op de rangschikking van de landen. In het DEA-RS model werden verschillende types gewichtsbeperkingen geformuleerd om de verhouding tussen verkeersdoden en zwaargewonden aan te geven. Dit leverde interessante resultaten op die ons inspireerden om dit model in de toekomst toe te passen op een zo uitgebreid mogelijke set van finale verkeersveiligheidsuitkomsten.

Met betrekking tot het benchmarken van het verkeersveiligheidsprogramma, dat gericht is op het vergelijken van de mens-voertuig-infrastructuurprestaties tussen landen en zo meer gedetailleerde aspecten van het ongevals- en verwondingsproces verklaren, werden verkeersveiligheidsindicatoren op het niveau van tussenliggende verkeersveiligheidsuitkomsten bestudeerd. De focus van dit onderzoek lag hierbij op de combinatie van individuele indicatoren in een samengestelde verkeersveiligheidsprestatie-index. Meer bepaald ontwikkelden we voor zes belangrijke verkeersveiligheidsfactoren (alcohol, snelheid, beschermende uitrusting, voertuig, weg, en medische hulpverlening) een uitgebreide set van hiërarchisch gestructureerde indicatoren om de verkeersveiligheidsprestatie van een land weer te geven. Hierbij werden diverse internationale gegevensbronnen geraadpleegd die indicatorwaarden verstrekken voor een grote reeks landen. In totaal werden 32 kwantitatieve prestatie-indicatoren gespecificeerd waarvoor gegevens verzameld (of berekend) werden voor 28 Europese landen, en de noodzakelijke gegevensverwerkingsprocedures (inclusief het detecteren van uitschieters en de imputatie van ontbrekende gegevens) werden uitgevoerd.

Om het multidimensionele concept van verkeersveiligheidsprestatie te vatten (hetgeen niet vastgelegd kan worden in één enkele indicator), onderzochten we of de DEA techniek gebruikt kan worden om een samengestelde verkeersveiligheidsprestatie-index te verkrijgen op basis waarvan landen met elkaar vergeleken kunnen worden. Hiervoor werd een meerlagig DEA-gebaseerd indexmodel (MLDEA-CI) opgesteld. Gebruikmakend van dit model werd de meest optimale verkeersveiligheidsprestatie-indexscore (berekend als combinatie van 32 hiërarchische prestatie-indicatoren) voor elk van de 28 Europese landen bepaald. De best presterende landen werden onderscheiden van de ondermaats presterende en landen werden gerangschikt. Een duidelijke link met het globale verkeersveiligheidsrisico (uit de benchmarking van het verkeersveiligheidsproduct) werd gevonden. Voorts werden land-specifieke voorbeeldlanden geïdentificeerd voor de ondermaats presterende landen en werd voor een land een goed inzicht verkregen in de domeinen waarin ondermaats gepresteerd werd door de indicatorgewichten te analyseren die in elke laag van de hiërarchie waren toegewezen. De resultaten geven zo een richting aan voor het verhogen van de verkeersveiligheidsprestatie in een land.

Bij de ontwikkeling van een samengestelde verkeersveiligheidsprestatie-index werd met het oog op een zinvolle en betrouwbare benchmarking bovendien onderzoek gedaan naar twee praktische uitdagingen op het gebied van gegevens (inclusief ontbrekende waarden en kwalitatieve indicatoren). Wat betreft de invloed van ontbrekende data in de dataset op de definitieve indexscore van de 28 Europese landen, werd gebruik gemaakt van intervallen, bepaald uit meervoudige imputatie, waarin de werkelijke waarden worden verondersteld te liggen. Een interval MLDEA-gebaseerd CI model werd later toegepast om voor ieder land een maximum en minimum indexscore te verkrijgen die respectievelijk overeenkomen met de meest gunstige en meest ongunstige optie. Het gebruik van een interval in plaats van de exacte indexscore voor elk land benadrukte de onderliggende imperfectie van de indicatorgegevens, en was een geloofwaardigere weergave van de globale verkeersveiligheidsprestatie van een land. Verder onderzochten we twee benaderingen binnen het DEA-domein om kwalitatieve (of ordinale) gegevens te modelleren in de context van een samengestelde index, met name het "imprecise DEA-based CI model" en het "fuzzy DEA-based CI model". Een enkele indexscore voor elk land werd verkregen door gebruik te maken van het op IDEA-gebaseerde CI model, hetgeen gemakkelijk te interpreteren en gebruiken is, terwijl in het op FDEA-gebaseerde CI model fuzzy indexscores verkregen werden die geschikt zijn om onzekerheden, eigen aan het menselijk denken, te vatten. De hoge mate van overeenstemming van het resultaat (de rangschikking) van deze twee modellen bewees hun robuustheid en impliceerde de mogelijkheid om één van beide benaderingen te gebruiken voor het modelleren van kwalitatieve gegevens.

Om te besluiten, is het internationale benchmarken van verkeersveiligheidsprestaties en de ontwikkeling hierin een veelbelovende stap om het verkeersveiligheidsniveau van een land te verbeteren. In dit proefschrift identificeerden we de belangrijkste onderzoeksuitdagingen met betrekking tot het benchmarken van het verkeersveiligheidsproduct en -programma, gebaseerd op verschillende types van verkeersveiligheidsindicatoren en ontwikkelden we de gepaste methodologie om deze uitdagingen te benaderen. Dit onderzoek droeg hoofdzakelijk bij tot de literatuur met betrekking tot het gebruik van de DEA-techniek en haar diverse uitbreidingen in het kader van zinvolle verkeersveiligheid benchmarkpraktijken. Hoewel het wiskundig van aard is, is de achterliggende theorie bevattelijk en kan het momenteel praktisch geïmplementeerd worden. Vanuit het standpunt van het verkeersveiligheidsbeleid en gebaseerd op de aanbevelingen uit de benchmark studies die betrekking hebben op het bepalen van doelstellingen en prioriteiten stellen aan acties, vormen het leren over goede praktijken die toegepast worden in voorbeeldlanden en het (opnieuw) formuleren van concrete veiligheidsstrategieën en -programma's bovendien de eerstvolgende te nemen stap voor elk land. Dit zal op haar beurt nieuwe uitdagingen en kansen creëren voor toekomstig onderzoek.