Soc Indic Res (2013) 114:739–756
DOI 10.1007/s11205-012-0171-0

Data Envelopment Analysis for Composite Indicators: A Multiple Layer Model

Yongjun Shen · Elke Hermans · Tom Brijs · Geert Wets

Accepted: 3 October 2012 / Published online: 18 October 2012

Abstract The development of a composite indicator (CI) over a set of individual indicators is worthwhile in case the methodological aggregation process is sound and the results are clear. It can then be used as a powerful tool for performance evaluation, benchmarking, and decision making. In this respect, data envelopment analysis (DEA), as a self appraisal technique, has recently received considerable attention in the construction of CIs for policy analysis and public communication. However, due to the ever increasing complexity of numerous performance evaluation problems, more and more potential indicators might be developed 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. Simply treating all the indicators to be in the same layer as is the case in the basic DEA model thereby ignores the information on their hierarchical structure, and further leads up to weak discriminating power and unrealistic weight allocations. To overcome this limitation, a multiple layer DEA-based CI model is developed in this study to embody a hierarchical structure of indicators in the DEA framework, and both its primal and dual form are realized. The proposed model is illustrated by constructing a composite road safety performance index for a set of European countries.

Keywords Composite indicators · Data envelopment analysis · Layered hierarchy · Road safety performance evaluation

Y. Shen (✉) · E. Hermans · T. Brijs · G. Wets
Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium
e-mail: yongjun.shen@uhasselt.be

E. Hermans
e-mail: elke.hermans@uhasselt.be

T. Brijs
e-mail: tom.brijs@uhasselt.be

G. Wets
e-mail: geert.wets@uhasselt.be

1 Introduction

During the last decades, a large number of composite indicators (CIs) or indexes 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, and been applied in wide ranging fields such as economy, society, governance, security, environment, sustainable development, globalization and innovation (Saisana and Tarantola 2002; Freudenberg 2003; Munda 2005; OECD 2008; Singh et al. 2009; Badea et al. 2011). According to a comprehensive review by Bandura (2008), around 180 different CIs have been identified all over the world. The proliferation of this kind of indexes is a clear symptom of their political importance and operational relevance in policy analysis and public communication. 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 (OECD 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 CI for cross country comparison includes both objective methods [e.g., principal component analysis (PCA), factor analysis (FA), regression analysis (RA), and neural networks (NNs)] and subjective methods [e.g., analytical hierarchy process (AHP), budget allocation (BA), and the technique for order preference by similarity to ideal solution (TOPSIS)] [see also Saisana and Tarantola (2002), OECD (2008), Shen et al. (2010a) 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 basis. 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 (DEA) (Charnes et al. 1978), originally developed as a non-parametric mathematical optimization technique to measure the so-called relative efficiency of a homogeneous set of decision making units (DMUs) 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 country 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 can be identified for each country separately, and policymakers 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 the road safety performance index (Hermans 2009).

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, Shen et al. (2010b; 2011) proposed a generalized multiple layer DEA (MLDEA) model (the primal form) to reflect the layered hierarchy of inputs and outputs by incorporating different types of possible weight restrictions for each category of each layer. In this paper, we explore the extension of this model for the construction of CIs, and further realize its dual form as well. The proposed model is illustrated by constructing a composite road safety performance index. The results are compared with those from the one layer DEA-based CI model, the equal weighting method, and the number of road fatalities per 10 billion passenger-kilometres travelled. The added value for policymakers is demonstrated subsequently.

The remainder of this paper is organized as follows. After a brief review of the basic DEA-based CI model in Sect. 2, we elaborate the extension of the model for hierarchical structure assessment in Sect. 3. In Sect. 4, we apply this multilayer model to combine the hierarchical road safety performance indicators into an overall index. The corresponding results are presented and discussed subsequently in Sect. 5. The paper ends with conclusion in Sect. 6.

2 DEA-based CI Model

Data envelopment analysis initially developed by Charnes et al. (1978) is a frontier analysis technique which employs linear programming tools to estimate the relations between multiple inputs and multiple outputs related to the DMUs under consideration. 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, which is presented as follows.

$$
\begin{array}{llll}
\text{Primal form:} & & \text{Dual form:} & \\
E_0 = \max \sum_{r=1}^{s} u_r y_{r0} & & \min & \theta_0 \\
s.t. & \sum_{i=1}^{m} v_i x_{i0} = 1, & s.t. & \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_0 x_{i0}, \quad i = 1,\ldots,m \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1,\ldots,n & & \sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{r0}, \quad r = 1,\ldots,s \\
& u_r, v_i \ge 0, \quad r = 1,\ldots,s, \quad i = 1,\ldots,m & & \lambda_j \ge 0, \quad j = 1,\ldots,n
\end{array}
\tag{1}
$$

The above linear programs (both primal and dual form) are computed separately for each DMU, and the subscript, $_0$, refers to the DMU whose relative efficiency is to be evaluated. y_{rj} and x_{ij} are the rth output and ith input respectively of the jth DMU, u_r is the weight given to the rth output, and v_i is the weight given to the ith input. λ_j $(j = 1,\ldots,n)$ is the dual weight given to the jth DMU's inputs and outputs, and θ_0 is the uniform proportional reduction in the DMU_0's inputs.

To use DEA for CI construction, i.e., aggregating a set of individual indicators into one overall index, however, 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 and 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 (DEA-CI) can be realized by converting the primal DEA model in (1) into the following constrained optimization problem, which is also known as the CCR model with constant inputs.

$$\begin{aligned} CI_0 &= \max \sum_{r=1}^{s} u_r y_{r0} \\ s.t. \quad & \sum_{r=1}^{s} u_r y_{rj} \leq 1, \quad j = 1, \ldots, n \\ & u_r \geq 0, \quad r = 1, \ldots, s \end{aligned} \tag{2}$$

The n DMUs 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) are all assigned with a value of unity. This linear program is run n times to identify the optimal index score for all DMUs by selecting their best possible indicator weights separately. In other words, the weights in the objective function are chosen automatically with the purpose of maximizing the value of DMU_0's composite index score and also respect the less than unity constraint for all the DMUs. Meanwhile, all the weights are required to be non-negative.[1] In general, a DMU is considered to be best-performing if it obtains an index score of one in (2), whereas a score less than one implies that it is underperforming.

Correspondingly, an equivalent dual form of the above problem can be formulated as follows, which can also be realized by assigning all the inputs with a value of unity in the dual form of the CCR model in (1).

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} \lambda_j \\ s.t. \quad & \sum_{j=1}^{n} y_{rj} \lambda_j \geq y_{r0}, \quad r = 1, \ldots, s \\ & \lambda_j \geq 0, \quad j = 1, \ldots, n \end{aligned} \tag{3}$$

3 Multiple Layer DEA-based CI Model

As a powerful performance measurement technique, the DEA-CI model has received significant attention in recent years due to its prominent advantages over other traditional methods as presented in Sect. 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

[1] This condition is normally replaced by using a small non-Archimedean number $\varepsilon > 0$ for restricting the model to assign a weight of zero to unfavorable indicators (Charnes and Cooper 1984).

DEA-CI model seems out of its capability to take this information into account. Moreover, simply putting all the indicators to be in the same layer may also lead up to weak discriminating power and unrealistic weight allocations of the model. Consequently, the development of a multiple layer DEA-based CI model (MLDEA-CI) is desirable, which will be elaborated in the following sections.

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 Fig. 1. $s^{(k)}$ is the number of categories in the kth layer ($k = 1, 2, \ldots, K$), and $s^{(1)} = s$. The development of the primal MLDEA-CI model can be considered as a two-step procedure. The first step is to 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.[2] With respect to the final layer, the second step is to determine the weights of all the sub-indexes by using the basic DEA-CI approach described in the previous section. Specifically, let $A^{(k)}_{f_k}$ 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:

$$y^{(K)}_{f_K 0} = \sum_{f_{K-1} \in A^{(K)}_{f_K}} w^{(K-1)}_{f_{K-1}} \left(\cdots \sum_{f_k \in A^{(k+1)}_{f_{k+1}}} w^{(k)}_{f_k} \left(\cdots \sum_{f_2 \in A^{(3)}_{f_3}} w^{(2)}_{f_2} \left(\sum_{f_1 \in A^{(2)}_{f_2}} w^{(1)}_{f_1} y^{(1)}_{f_1 0} \right) \right) \right)$$
$$\sum_{f_k \in A^{(k+1)}_{f_{k+1}}} w^{(k)}_{f_k} = 1, \; w^{(k)}_{f_k} \geq 0, \; f_k = 1, \ldots, s^{(k)}, \; k = 1, \ldots, K-1 \tag{4}$$

where $w^{(k)}_{f_k}$ denotes the internal weights associated with the indicators of the fth category in the kth layer, which are non-negative[3] 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 $y^{(K)}_{f_K 0}$ from (4) to model (2), we obtain the following objective function:

$$CI_0 = \max \sum_{f_K=1}^{s^{(K)}} u_{f_K} \left(\sum_{f_{K-1} \in A^{(K)}_{f_K}} w^{(K-1)}_{f_{K-1}} \left(\cdots \sum_{f_k \in A^{(k+1)}_{f_{k+1}}} w^{(k)}_{f_k} \left(\cdots \sum_{f_2 \in A^{(3)}_{f_3}} w^{(2)}_{f_2} \left(\sum_{f_1 \in A^{(2)}_{f_2}} w^{(1)}_{f_1} y_{f_1 0} \right) \right) \right) \right) \tag{5}$$

where u_{f_K} is the weight given to the fth category in the Kth layer (i.e., the final layer), $f_K = 1, \ldots, s^{(K)}$.

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:

---

2 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.

3 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.

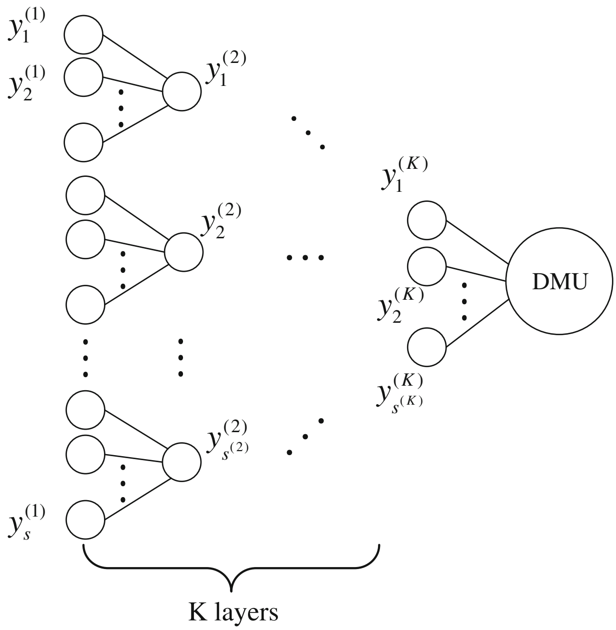

Fig. 1 A hierarchical structure of indicators

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

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

$$\begin{aligned}
\sum_{f_1 \in A^{(2)}_{f_2}} \hat{u}_{f_1} &= \prod_{k=2}^{K-1} \underset{f_k \in A^{(k+1)}_{f_{k+1}}}{w^{(k)}_{f_k}} \cdot u_{f_K} \\
&\vdots \\
\sum_{f_1 \in A^{(K-1)}_{f_{K-1}}} \hat{u}_{f_1} &= \underset{f_{K-1} \in A^{(K)}_{f_K}}{w^{(K-1)}_{f_{K-1}}} u_{f_K} \\
\sum_{f_1 \in A^{(K)}_{f_K}} \hat{u}_{f_1} &= u_{f_K}
\end{aligned} \tag{7}$$

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

$$\underset{f_k \in A^{(k+1)}_{f_{k+1}}}{w^{(k)}_{f_k}} = \frac{\sum_{f_1 \in A^{(k)}_{f_k}} \hat{u}_{f_1}}{\sum_{f_1 \in A^{(k+1)}_{f_{k+1}}} \hat{u}_{f_1}}, \quad f_k = 1, \ldots, s^{(k)},\ k = 1, \ldots, K-1 \tag{8}$$

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. There are a variety of weight restriction techniques proposed in the DEA literature, such as absolute weight restrictions, relative weight restrictions, ordinal weight restrictions, and virtual weight restrictions (Wong and Beasley 1990; Allen et al. 1997; Thanassoulis et al. 2004).

Now, by incorporating the deduced internal weights and appropriate weight restrictions into model (2), we obtain the primal MLDEA-CI model as follows:

$$
\begin{aligned}
& CI_0 = \max \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 0} \\
& s.t. \quad \sum_{f_1=1}^{s} \hat{u}_{f_1} y_{f_1 j} \leq 1, \quad j = 1, \ldots, n \\
& \qquad \sum_{f_1 \in A_{f_k}^{(k)}} \hat{u}_{f_1} \Bigg/ \sum_{f_1 \in A_{f_{k+1}}^{(k+1)}} \hat{u}_{f_1} = \underset{f_k \in A_{f_{k+1}}^{(k+1)}}{w_{f_k}^{(k)}} \in \Theta, \quad f_k = 1, \ldots, s^{(k)},\ k = 1, \ldots, K-1 \\
& \qquad \hat{u}_{f_1} \geq 0, \quad f_1 = 1, \ldots, s
\end{aligned}
\tag{9}
$$

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 acceptable layer-specific weights are guaranteed, which cannot be realized in the one layer model.

3.2 The Dual MLDEA-CI Model

Having developed the primal MLDEA-CI model, we now further deduce its equivalent dual form. In doing so, we firstly generalize the vector form of the primal model (9) as follows:

$$
\begin{aligned}
CI_0 = \max \quad & \hat{u}\boldsymbol{y}_0 \\
s.t. \quad & \hat{u}\boldsymbol{Y} \leq 1 \\
& \hat{u}\boldsymbol{Q} \leq 0 \\
& \hat{u} \geq 0
\end{aligned}
\tag{10}
$$

where $\boldsymbol{Q}$ is a $s \times t$ weight restriction matrix corresponding to the second constraint in model (9), in which t denotes the total number of weight restriction.[4] Consequently, the dual MLDEA-CI model can be expressed as follows:

$$
\begin{aligned}
\min \quad & e_n \boldsymbol{\lambda} \\
s.t. \quad & \boldsymbol{Y\lambda} + \boldsymbol{Q\tau} \geq \boldsymbol{y}_0 \\
& \boldsymbol{\lambda} \geq 0,\ \boldsymbol{\tau} \geq 0
\end{aligned}
\tag{11}
$$

where e_n is a row vector $(1 \times n)$ with all elements equal to one, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \ldots \lambda_{\boldsymbol{n}})^{\boldsymbol{T}}$ is the dual weight vector with the same definition as in model (3), and $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots \tau_{\boldsymbol{m}})^{\boldsymbol{T}}$ is an extra vector due to the incorporation of weight restrictions in model (10).

[4] As has been proved in Shen et al. (2011), no matter which weight restriction(s) is imposed, it will maintain the model to be linear. As a result, the model (10) has the similar vector form as the assurance region (AR) model (Wong and Beasley 1990), except for the definition of the weight restriction matrix $\boldsymbol{Q}$, in which only two elements of each column are allocated with non-zero values in the AR model since merely pairwise comparisons of the weights can be carried out. In the MLDEA-CI model, however, comparisons among different types of weight combinations can be realized.

4 Construction of a Road Safety Performance Index

To demonstrate the effectiveness of the proposed MLDEA-CI model, we apply it (both primal and dual form) to construct a composite road safety performance index for a set of European countries. As we know, road traffic crashes and consequent injuries and fatalities have been recently recognized as one of the most important public health issues that requires concerted efforts for effective and sustainable prevention (WHO 2004, 2009). To make progress on road safety, rather than only focusing on crash data, more attention has nowadays been paid to the underlying risk factors influencing safety, such as road user behavior, which is recognized as one of the largest contributors to road crashes with a sole cause in 57 % of all crashes and a contributing factor in over 90 % (Treat et al. 1977; Rumar 1982; Green and Senders 2004). In this respect, a number of relevant safety performance indicators (SPIs) are developed in order to measure the performance on different risk aspects of road user behavior and to understand the process that leads to crashes. In this study, the proposed MLDEA-CI model is applied to combine these SPIs into one overall index. In doing so, the set of indicators and its hierarchical structure, the indicator data, as well as the weight restrictions are three main aspects that need to be specified before applying the model.

4.1 Indicators on Road User Behavior and Their Hierarchy

In Europe, the most commonly used SPIs in terms of human behavior for road transport are the incidence of drinking and driving, speed measurements, and the use of various protective systems (Hakkert et al. 2007). For each of these three risk factors, several SPIs are developed constituting a multilayer hierarchical structure. More specifically, concerning drink driving, *the percentage of drivers above legal alcohol limit in roadside police tests* and *the percentage of alcohol-related fatalities* are the two representative indicators. Moreover, to describe driving at inappropriate or excessive speed which has been widely accepted as a highly important factor influencing both the number of crashes and the severity of injuries, *the mean speed* and the level of compliance of vehicles (or *the proportion of vehicles exceeding the speed limit*) in free-flowing traffic are 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 third behavioral characteristic that is believed to influence road safety is the use of various protective systems such as seat belts and child restraints. As indicators for this risk aspect, *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. Totally, 11 hierarchically structured SPIs are specified reflecting the performance of road user behavior in each country, which are shown in Fig. 2.

4.2 Data Collection and Processing

After the formulation of SPIs describing the characteristics of the road user behavior, we now collect available indicator data for different countries from a wide range of international data sources, such as the European Transport Safety Council (2010) and Vis and Eksler (2008). In this study, average values of 2006–2008 are obtained for the 13 European countries[5] being Austria (AT), Belgium (BE), Finland (FI), France (FR), Hungary (HU),

[5] Missing data are imputed by using Multiple Imputation in SPSS 20.0 (IBM Corp. 2011).

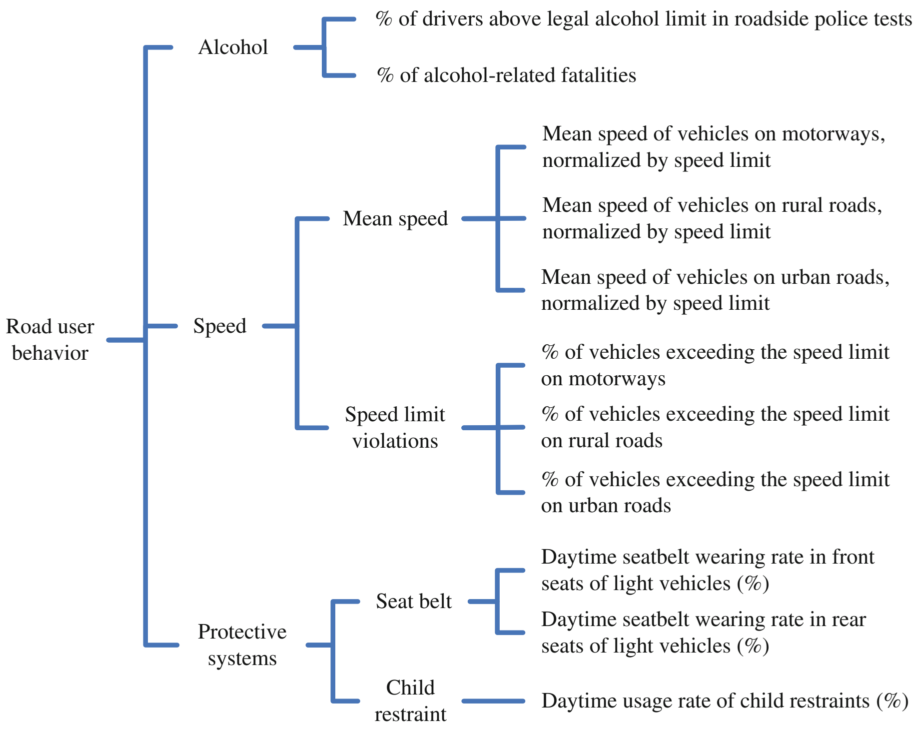

Fig. 2 The hierarchical framework of SPIs with respect to road user behavior

Ireland (IE), Lithuania (LT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SI), Sweden (SE), and Switzerland (CH).

Prior to the application of the MLDEA-CI model, the raw data have to be first normalized so as to eliminate the scale differences of the indicators, 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 (OECD 2008) is adopted because the ratio of two numbers is best kept by this approach:

$$\tilde{y}_{rj} = \begin{cases} y_{rj} \big/ y_j^*, \forall_j, y_j \textit{ is a benefit indicator} \\ y_j^- \big/ y_{rj}, \forall_j, y_j \textit{ is a } \text{cost } \textit{indicator} \end{cases} \tag{12}$$

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 for normalization when a benefit respectively a cost indicator is taken into account. In this study, all the eight alcohol and speed SPIs are identified as cost indicators, while the three SPIs related to protective systems are benefit ones. As a result, the country with the highest safety performance (or the lowest human behavior risk) receives a normalized value of one whereas the others are expressed as percentage share of that country's value. The resulting normalized data (13 × 11) based on (12) are presented in Table 1. Taking the percentage of alcohol-related fatalities (*I2*) as an example, the Netherlands performs the best (1.000) while Slovenia worst (0.078), and all other countries' values lie within this interval.

4.3 Weight Restrictions

In addition to the selection of indicators, as well as the data collection and processing, 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, as well as the two aspects related to speed. Thus, we obligate the weights to vary within a range from 0.8 to 1.2 of their average weights. Taking the two alcohol indicators as an example, their weights are thereby required to lie between 0.4 and 0.6. With regard to the last layer, to guarantee that all the three risk factors of road user behavior, i.e., alcohol, speed, and protective systems, will be used to some extent by the model, the share[6] of each of these three risk factors 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.

5 Results

The proposed MLDEA-CI model can now be applied to compute the composite road safety index score for the 13 European countries with respect to their road user behavior. Best-performing countries are therefore distinguished from underperforming ones and countries ranked. Moreover, useful benchmarks for the underperforming countries can be identified by using the dual model (11), and the indicator weights allocated in each layer of the hierarchy can be deduced for each country based on the primal model (9). All the model outputs are illustrated in the following sections.

5.1 Index Score and Country Ranking

With the developed MLDEA-CI model (9), the 11 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 2, along with the ones from the basic DEA-CI model (2).

Since a large number of indicators (11) relative to the number of countries (13) is considered in this study,[7] most of the countries 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 user behavior. 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 three countries: the Netherlands, Sweden, and Switzerland. With respect to the remaining

[6] The share is the sum of the products of the indicator values and the corresponding weights divided by the final index score.

[7] A rough rule of thumb for applying the DEA model is to choose the number of DMUs, i.e., n, equal to or greater than $max\{m \times s, 3 \times (m + s)\}$, where m and s represent the number of inputs and outputs, respectively (Cooper et al. 2000), which in this study should be 36 at least.

Table 1 Normalized data on the 11 hierarchical SPIs

	Alcohol		Speed						Protective systems		
			Mean speed			Speed limit violation (%)			Seat belt		Child restraint
	% of drivers above legal alcohol limit in roadside police tests	% of alcohol-related fatalities	Motor-ways	Rural roads	Urban roads	Motor-ways	Rural roads	Urban roads	Daytime seatbelt wearing rate in front seats of light vehicles (%)	Daytime seatbelt wearing rate in rear seats of light vehicles (%)	Daytime usage rate of child restraints (%)
	I1	*I2*	*I3*	I4	*I5*	*I6*	*I7*	*I8*	*I9*	*I10*	*I11*
AT	0.116	0.463	0.917	0.781	0.802	0.766	0.051	0.254	0.904	0.669	0.863
BE	0.068	0.654	0.827	0.743	0.768	0.348	0.029	0.222	0.799	0.467	0.729
FI	0.593	0.136	0.942	0.729	0.907	0.409	0.023	0.323	0.911	0.949	0.716
FR	0.263	0.123	0.912	0.787	0.838	0.505	0.037	0.318	1.000	0.957	0.937
HU	0.279	0.283	0.973	0.793	0.817	0.362	0.033	0.230	0.727	0.479	0.433
IE	0.237	0.119	0.924	0.762	0.724	1.000	0.032	0.223	0.901	0.875	0.857
LT	0.555	0.321	0.978	0.713	0.714	0.789	0.025	0.318	0.609	0.350	0.404
NL	0.081	1.000	0.879	0.740	0.881	0.454	0.020	0.234	0.959	0.852	0.758
PL	0.091	0.438	0.788	0.697	0.647	0.290	0.015	0.165	0.799	0.564	0.905
PT	0.137	0.610	0.828	0.618	0.919	0.302	0.014	0.360	0.881	0.549	0.591
SI	0.122	0.078	0.943	1.000	0.713	0.480	1.000	0.163	0.874	0.527	0.672
SE	1.000	0.357	0.864	0.717	0.870	0.241	0.019	0.259	0.973	0.887	1.000
CH	0.277	0.230	0.922	0.757	1.000	0.710	0.043	1.000	0.887	0.770	0.895

Table 2 Index score of the 13 European countries based on the basic DEA-CI model and the MLDEA-CI model

	Basic DEA-CI	MLDEA-CI
NL	1.000	1.000
SE	1.000	1.000
CH	1.000	1.000
AT	1.000	0.936
FI	1.000	0.926
FR	1.000	0.865
IE	1.000	0.864
LT	1.000	0.846
PT	0.978	0.835
BE	0.938	0.829
PL	0.955	0.806
HU	1.000	0.728
SI	1.000	0.679

countries, a relatively lower index score is obtained compared to that from the basic DEA-CI model, and they are all underperforming. The same results can be derived using the dual MLDEA-CI model (11) as well.

Apart from identifying the best-performing and underperforming countries, we further rank all these countries by computing their cross-index score, which reflects their all round road safety performance by taking the best possible weights for each country in the data set into account. For more information on this technique, we refer to Shen et al. (2012). The results and the corresponding ranking are presented in Table 3, together with the ones from applying the equal weighting approach, and the number of road fatalities per 10 billion passenger-kilometres (pkm) travelled. The comparative results are also illustrated in Fig. 3.

Comparing the index scores based on the MLDEA-CI and the simple equal weighting in Fig. 3, we can see that a relatively higher value is derived for each country by using the former method, which verifies the fact that the DEA approach puts each country in its most favorable light, and any other weighting scheme would generate a lower composite score. By checking the two ranking results in Table 3, we find that they are highly correlated with the correlation coefficient of 0.951, and all the countries have a difference in rank with a maximum of two positions. Moreover, as a relevant point of reference, the number of road fatalities per 10 billion passenger-kilometres travelled is computed for these 13 countries based on the average values of 2006–2008. The high degree of consistency between the index rankings and fatality ranking can be observed, especially for those best-performing (e.g., Sweden, the Netherland, and Switzerland) and worst-performing countries (e.g., Hungary, Poland, and Lithuania), and the index score generated from the MLDEA-CI model produces even a higher negative correlation with the number of fatalities than that from the equal weighting method (−0.813 vs. −0.786), which indicates the effectiveness of the proposed model, and further implies that the created road safety performance index has a clear link with road safety outcome, and can be used as a valuable predictor based on which efficient policy measures can be put forward.

5.2 Benchmarks from the Dual Model

To better understand the computational process leading to the index score presented in the last column of Table 2, especially the reasons why the 10 underperforming countries are

Table 3 Country rankings based on MLDEA-CI, equal weighting, and the number of road fatalities per 10 billion pkm travelled

	Cross-index based on MLDEA-CI	Ranking	Index based on equal weighting	Ranking	Number of road fatalities per 10 billion pkm*	Ranking
SE	1.000	1	0.713	1	45	1
NL	0.963	2	0.636	2	48	3
CH	0.930	3	0.618	3	45	2
AT	0.863	4	0.570	6	97	8
FI	0.844	5	0.581	4	56	4
IE	0.829	6	0.554	7	70	6
FR	0.828	7	0.572	5	63	5
SI	0.784	8	0.501	10	107	9
PT	0.775	9	0.511	8	109	10
BE	0.774	10	0.510	9	93	7
LT	0.769	11	0.490	12	171	11
PL	0.725	12	0.497	11	224	12
HU	0.687	13	0.444	13	271	13

* Average value of 2006–2008 (European Commission 2012)

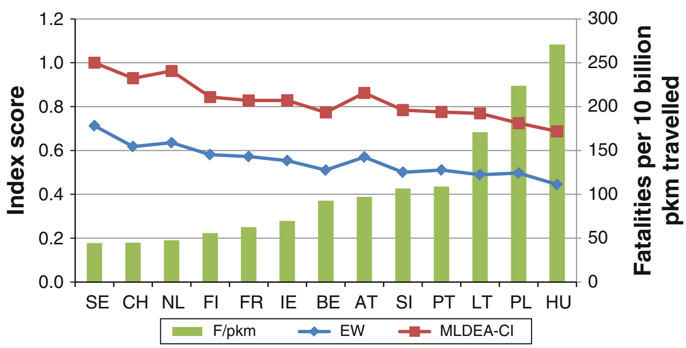

Fig. 3 Index scores based on MLDEA-CI, equal weighting, and the number of road fatalities per 10 billion pkm travelled

unable to obtain a value of one, we further explore the mechanism of the MLDEA-CI model.

Theoretically, the primal MLDEA-CI model (9) 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 a higher road safety performance than country *A* with respect to at least one risk aspect of the road user behavior since the index score of *B* is relatively higher with the same set of weights. Therefore, country *B* can be viewed as a valuable benchmark for country *A*, and a series of benchmark countries like *B* constitute a reference set for country *A* to learn from.

Moreover, from the view of the dual MLDEA-CI model (11), the dual weights, i.e., λ, can be seen as indicating the amount of technical weight that is attributed by each benchmark country in the construction of a hypothetical composite unit. In other words, the countries with non-zero dual weights make up the reference set for the country under study, and the value of dual weight also indicates the relative importance of a benchmark country within the reference set. Using this principle, the reference sets and dual weights of all the 10 underperforming countries are illustrated in Table 4, and the benchmark country with the greatest dual weight is highlighted.

Taking Austria as an example, the best possible indicator weights assigned for this country only result in its optimal index score of 0.936 because a weighted score of one is achieved by two other countries, which are Sweden and Switzerland. Therefore, Austria is underperforming, and it could take a hypothetical composite country that consists of the above two best-performing countries as an example, in which Switzerland appears to be the more important benchmark as it obtains a greater dual weight (0.538) than Sweden (0.397). In other words, Austria should learn the most from Switzerland to improve its road user behavior.

5.3 Detailed Weights Allocation in the Primal Model

In addition to the identification of specific benchmarks for all the underperforming countries, valuable insight can also be gained by exploring the indicator weights allocated in each layer of the hierarchy per country from the view of the primal MLDEA-CI model (9). More specifically, in the basic DEA-CI model (2), 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, while low weights provide policymakers with valuable information about the aspects requiring most action for improvement. In Fig. 4, the assigned weights (the values in brackets are shares) based on the primal MLDEA-CI model (9) 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 (0.936) in the multilayer model.

Table 4 Benchmarks for the underperforming countries

	NL	SE	CH
AT		0.397	**0.538**
BE	0.014	**0.719**	0.095
FI		**0.746**	0.180
FR	0.046	**0.600**	0.219
HU		**0.666**	0.062
IE	0.114	**0.749**	
LT	0.159	**0.686**	
PL		**0.707**	0.099
PT	0.157	**0.570**	0.107
SI		**0.679**	

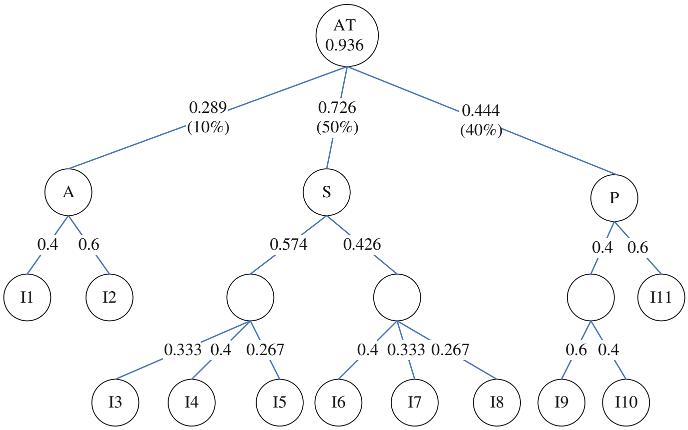

Fig. 4 Assigned weights (and shares) in each layer of the hierarchy for Austria based on the MLDEA-CI model

Figure 4 shows the accordance of the weights (and shares) with the imposed restrictions described in Sect. 4.3. For instance, the indicators belonging to a particular category of each layer, such as the three mean speed indicators (*I3–I5*), are of similar importance (with a 20 % variability of their average weights), and the shares of the three risk factors of road user behavior are all within the range of 10-50 %. 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 by the same principle, we find that Austria is doing relatively well in the speed aspect (with the highest share of 50 %), especially the mean speed on rural roads (*I4*). Whereas more policy attention should be paid to the risk aspect on alcohol (with the lowest share of 10 %), followed by the protective systems (with a share of 40 %), in which improving rear seat belt wearing rate (*I10*) is most urgently needed.

To conclude, although a relatively lower index score is achieved, the MLDEA-CI model is to be preferred since it produces valuable results by means of taking all the indicators and their hierarchical structure into account, which, however, cannot be (correctly) offered from the basic DEA-CI model.

6 Conclusion

Due to the remarkable ability of integrating large amounts of information into understandable formats that are often easier to interpret than finding a common trend in many separate indicators, composite indicators or indexes are increasingly recognized as a useful tool in policy analysis and public communication. In this study, we investigated the use of data envelopment analysis for composite index construction. Starting from the basic DEA-based CI model, we further explored the incorporation of a layered hierarchy of indicators in the DEA framework, and developed a multiple layer DEA-based CI 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. Moreover, value judgment from decision makers or experts can be easily incorporated by restricting the weight flexibility in a particular category of a layer. In doing so, the obtainment of realistic and acceptable indicator weights is guaranteed, and it cannot be realized in the one layer model. In the application, the proposed MLDEA-CI model was applied to construct a composite road safety performance index for 13 European countries by combining 11 indicators with respect to their road user behavior. Due to the consideration of hierarchical information on the indicators and the incorporation of corresponding weight restrictions on the different layers, the discriminating power of the model was obviously improved compared to the basic DEA-CI model, and three best-performing countries were distinguished from underperforming ones. Moreover, the ranking of these countries was deduced by computing their cross-index score. A clear link with the road safety outcome, i.e., the number of road fatalities per passenger-kilometres travelled, was verified, and a higher correlation coefficient was found by using the MLDEA-CI model than the equal weighting method. Furthermore, rather than postulating the same set of benchmarks 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, which however, cannot be correctly offered from either the one layer model or the equal weighting method. In addition, since road user behavior is only one of the main risk factors contributing to road crashes, other factors such as the vehicle compatibility, road design, and post-crash medical treatment should also be taken into account in the future to evaluate the all round road safety performance of a country. Such a complex multilayer hierarchical picture makes the MLDEA-CI model particularly valuable.

References

Adolphson, D. L., Cornia, G. C., & Walters, L. C. (1991). *A unified framework for classifying DEA models, In: Operational research '90* (pp. 647–657). New York: Pergamon Press.

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,73*, 13–34.

Badea, A. C., Rocco S, C. M., Tarantola, S., & TaBolado, R. (2011). Composite indicators for security of energy supply using ordered weighted averaging. *Reliability Engineering & System Safety,96*(6), 651–662.

Bandura, R. (2008). A survey of composite indices measuring country performance: 2008 update, United Nations Development Programme—Office of Development Studies.

Bao, Q., Ruan, D., Shen, Y., Hermans, E., & Janssens, D. (2012). Improved hierarchical fuzzy TOPSIS for road safety performance evaluation. *Knowledge-Based Systems,32*, 84–90.

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,15*(3), 333–334.

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

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

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

Cherchye, L., Moesen, W., Rogge, N., van Puyenbroeck, T., Saisana, M., Saltelli, A., et al. (2008). Creating composite indicators with DEA and robustness analysis: The case of the technology achievement index. *Journal of the Operational Research Society,59*, 239–251.

Cooper, W. W., Seiford, L. M., & Tone, K. (2000). *DEA: A comprehensive text with models, application, references and DEA solver software*. Boston: Kluwer Academic Publishers.

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

European Commission (EC). (2012). *EU energy and transport in figures 2012*. Brussels: Commission of the European Communities.

European Transport Safety Council (ETSC). (2010). Road safety target in sight: Making up for lost time, 4th road safety PIN report, ETSC, Brussels.

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

Freudenberg, M. (2003). Composite indicators of country performance: A critical assessment, STI working paper 2003/16, Organization for Economic Co-operation and Development, Paris.

Green, M., & Senders J. (2004). Human error in road accidents, visual expert, Canada. http://www.visualexpert.com/Resources/roadaccidents.html.

Hakkert, A. S., Gitelman, V., & Vis, M. A. (Eds.). (2007). Road safety performance indicators: Theory. Deliverable D3.6 of the EU FP6 project SafetyNet.

Hermans, E. (2009). A methodology for developing a composite road safety performance index for cross-country comparison. PhD thesis, Hasselt University, Hasselt, Belgium.

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.

Munda, G. (2005). Multi-criteria decision analysis and sustainable development. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple-criteria decision analysis: State of the art surveys* (pp. 953–986). New York: Springer International Series in Operations Research and Management Science.

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.

Ramanathan, R. (2006). Evaluating the comparative performance of countries of the Middle East and North Africa: A DEA application. *Socio-Economic Planning Sciences,40*, 156–167.

Rumar, K., (1982). The human factor in road safety, invited paper at 11th Australian Road Research Board Conference, Melbourne.

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.

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 and Prevention,48*, 430–441.

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,38*(12), 15262–15272.

Shen, Y., Li, T., Hermans, E., Ruan, D., Wets, G., Vanhoof, K., et al. (2010a). A hybrid system of neural networks and rough sets for road safety performance indicators. *Soft Computing,14*(12), 1255–1263.

Shen, Y., Hermans, E., Ruan, D., Wets, G., Vanhoof, K., Brijs, T. (2010b). A generalized multiple layer data envelopment analysis model. In: *Proceedings of the 4th international conference on intelligent systems and knowledge engineering (ISKE'09)* (vol. 2, pp. 215–220). World Scientific Proceedings Series on Computer Engineering and Information Science: Diepenbeek

Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators,9*, 189–212.

IBM Corp. (2011). IBM SPSS missing values 20. http://www.ibm.com/spss.

Thanassoulis, E., Portela, M. C., & Allen, R. (2004). Incorporating value judgements in DEA. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.), *Handbook on data envelopment analysis* (pp. 99–138). Boston: Kluwer Academic Publishers.

Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., Hume, R.D., Mayer, R.E., et al. (1977). Tri-level Study of the causes of traffic accidents, report No. DOT-HS-034-3-535-77 (TAC).

Vis, M. A. & Eksler, V. (Eds.). (2008). Road safety performance indicators: Updated country comparisons, deliverable D3.11a of the EU FP6 project SafetyNet.

Wong, Y. H. B., & Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society,41*(9), 829–835.

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). (2009). Global status report on road safety—time for action, WHO, Geneva. http://www.who.int/violence_injury_prevention/road_ safety_status/2009/en/.

Zhou, P., Ang, B. W., & Poh, K. L. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics,62*, 291–297.