Transportation Research Interdisciplinary Perspectives 29 (2025) 101326

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.elsevier.com/locate/trip

Assessing the sustainable performance of the transport sector in European countries using alternative Benefit-of-the-Doubt models

Sarah B. Gruetzmacher [a], Clara B. Vaz [b,*], Ângela P. Ferreira [b]

[a] *Department of Electrical Engineering, Federal University of Technology - Paraná (UTFPR), Curitiba, 80230-901, Brazil*
[b] *Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, 5300-253, Portugal*

ARTICLE INFO

Dataset link: https://ec.europa.eu/eurostat/data/database

Keywords:
Transport sector
Sustainable performance
European countries
Data envelopment analysis
Benefit-of-the-Doubt models

ABSTRACT

The energy policy of the European Union stresses the need for sustainable energy consumption, improvements in energy efficiency and lower fossil fuel dependence in a decoupling strategy from unstable democracies. Transportation still represents a sector largely dependent on fossil fuels, which come with several negative impacts. Measuring and assessing the sustainability of the transport sector becomes necessary. This study aims to assess the sustainability performance of the transport sector across 28 European countries over a four-year period, aligned with the policy agenda outlined in strategic documents. The methodological approach involves applying Benefit-of-the-Doubt (BoD) models, comparing a version that uses transformation methods for anti-isotonic sub-indicators with a variant that directly incorporates these sub-indicators as reverse indicators. In general, the European countries have improved the sustainability performance of their transport sector during the time span analyzed according to the results of both models. For the inefficient units, two improvement strategies are presented based on the profiles identified on the benchmarks from both models, which can be alternative stages to achieve the robust best practices of the benchmarks.

1. Introduction

The increased awareness of the impact of human activities on the environment, with economic, social, and environmental costs, emphasizes the growing interest in sustainable development. In the framework of the Sustainable Development Goals established by the United Nations General Assembly in 2015 (United Nations, 2022) and Europe's current strategic actions and challenges, transport is one of the sectors with major potential to contribute to a low-carbon economy and an energy-efficient framework. In fact, transportation is the only sector where greenhouse gas (GHG) emissions have been steadily increasing (European Environment Agency, 2021). In 2019, the transport sector accounted for 30.2% of total energy-related GHG emissions (European Commission, 2021b). Between 2010 and 2050, passenger transport in the European Union (EU) is projected to increase by 42%, while freight transport is estimated to increase by 60% (European Commission, 2019b). Reducing socioeconomic and environmental pressures in the transport sector is, therefore, crucial to achieving long-term sustainability (European Environment Agency, 2021).

The EU follows several transport-related European strategic documents, namely the 2011 White Paper on transport - Roadmap to a Single European Transport Area (European Commission, 2011), which identified the main challenges for the development of the EU transport sector and proposed strategies for deep changes in the sector aiming at a more sustainable and efficient system: the proposals will dramatically reduce Europe's dependence on imported oil and cut carbon emissions in transport by 60% by 2050. The European Commission continues to develop and implement new strategies and initiatives to address emerging challenges and priorities in the transport sector, as for example, the Sustainable and Smart Mobility Strategy (European Commission, 2020), the Fit for 55 package with new targets for emissions reduction in the transport sector (European Commission, 2021a), and horizon scanning methodologies as proposed by Tsakalidis et al. (2021) to identify future trends and challenges in transport research and innovation governance within the European context. These efforts aim to build upon the foundations laid by the 2011 White Paper, ensuring that Europe's transport policies remain effective and moving forward. With regard to the Sustainable Development Goals from the United Nations (UN), the ones related to the transport sector were highlighted in Gudmundsson and Regmi (2017). These strategic documents depict that the sustainable development of the transport sector has been put

* Corresponding author at: Research Centre in Digitalization and Intelligent Robotics (CeDRI), Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Bragança, 5300-253, Portugal.
E-mail addresses: sarah@alunos.utfpr.edu.br (S.B. Gruetzmacher), clvaz@ipb.pt (C.B. Vaz), apf@ipb.pt (Â.P. Ferreira).

https://doi.org/10.1016/j.trip.2025.101326
Received 20 September 2024; Received in revised form 30 December 2024; Accepted 2 January 2025
Available online 18 January 2025

on the agenda of European countries, highlighting the necessity of measuring and assessing the current sustainable transport performance in order to achieve these targets. They also emphasize the importance of analyzing sustainable transport planning, since transport policy and planning decisions can have diverse and long-term impacts on sustainable development (United Nations Department of Economic and Social Affairs, 2011).

The sustainability assessment of the transport sector becomes more relevant as a key input for decision-making and monitoring the progress towards the existing targets. Sub-indicators can be valuable tools to create a process to assist in identifying the best practices in European countries for transport-related stakeholders, as they are generally understood to be able to measure the performance of decision-making units (DMU) in a specific area. However, a single sub-indicator cannot completely describe multi-faceted subjects like environment, economy, society and/or technological development. Besides, the analysis of several sub-indicators is complex for decision-makers, as a joint interpretation can be difficult (Joumard and Gudmundsson, 2010). The best approach is to aggregate the individual sub-indicators into a single index, the composite indicator (CI), to measure and assess the performance that individual sub-indicators cannot completely capture (Reisi et al., 2014).

Different models can be used to aggregate the various sub-indicators into the CI score. The traditional CI has some limitations, such as normalization challenges as different techniques may affect the results, and the requirement for pre-defined weights, which introduces subjectivity, among others. Additionally, comparing each CI calculation with the mean score helps facilitate the interpretation of multiple scores. The CI derived from Data Envelopment Analysis (DEA) models overcomes these drawbacks since they endogenously defines the weights, providing a more objective and data-driven approach.

Currently, one of the widely employed DEA based models for constructing composite indicators is the Benefit-of-the-Doubt (BoD) model. The weighting procedure based on the "benefit of doubt" principle was originally proposed by Melyn and Moesen (1991) in the context of macroeconomic performance evaluation. This model can only derive the CI by aggregating isotonic sub-indicators (Dyson et al., 2001), capturing the positive aspects of performance, *i.e.*, their increased values increase the performance. However, performance assessment using CI frequently has to handle anti-isotonic sub-indicators, where their increased values deteriorate the performance, incorporating, in this way, the performance's negative aspects. Many useful and important sub-indicators, for instance, air pollution and traffic accidents, belong to this category. One approach to managing anti-isotonic sub-indicators is to use data transformation methods (Dyson et al., 2001; Scheel, 2001), by which they can be transformed into isotonic sub-indicators and incorporated into the BoD model. Another alternative is to use the directional distance function (DDF) model, which allows the simultaneous radial reduction of anti-isotonic sub-indicators and increase of isotonic sub-indicators in order to reach the best practice frontier (Zanella et al., 2015a,b; Charles et al., 2016; Elisa Fusco, 2015; Vidoli et al., 2015). Recently, Färe et al. (2019) proposed a model that can directly aggregate the anti-isotonic sub-indicators into the CI without the need for any further transformation by treating them as reverse rather than undesirable sub-indicators. This study provides two significant contributions. It focuses on evaluating the performance of the transport sector at the European level, identifying the best practices countries over a time frame in which the short to medium-term actions outlined in the 2011 White Paper are already envisioned and able to add value in guiding the development of EU transport policy to achieve long-term objectives.

The performance assessment of the transport sector towards sustainable mobility in the European Union (EU) is performed over a time span of four years, prior to the Covid-19 pandemic, to avoid data biases introduced by the decreased activity. The sub-indicators to assess the performance of the transport sector in EU countries were chosen considering the targets for sustainable mobility set in the 2011 White Paper on transport from the European Commission (2011) and the Sustainable Development Goals adopted by the United Nations (United Nations, 2022), as discussed in Gruetzmacher (2021) and Gruetzmacher et al. (2021a). Additionally, this study makes a theoretical contribution by comparing the Benefit-of-the-Doubt model using transformation methods for the anti-isotonic sub-indicators and the variant of the BoD model proposed by Färe et al. (2019). In order to fulfill this objective, an illustrative example, prior to the case study application, is used to explore the differences in the performance assessment derived from several approaches.

This paper is structured as follows: the second section presents an overview of recent works on sustainability analysis at the European level and a comprehensive literature review on the methodological approaches for sustainability measurement in the transport sector; Section 3 describes the methodological approach, presenting the two alternative Benefit-of-the-Doubt (BoD) models to derive the composite indicators and presents a small example able to explore their performance and main differences; Section 4 details the case study and describes the sub-indicators selected to compute the CI as well as their descriptive analysis. Section 5 presents and discusses the results from the CI obtained. Finally, the conclusions from this work are rounded up in Section 6.

2. Literature review

2.1. Sustainability performance in the European transport sector

The analysis of the sustainability performance of the transport sector in European countries yields several important policy implications and strategies to achieve sustainability targets under the priorities and goals established at the European and United Nations General Assembly levels. A systematic literature review on road freight transportation (Nkesah, 2023) highlighted the importance of policy and regulation, efficient logistics and innovative technology-based measures in its sustainability. Substantial research has been conducted on sustainable performance assessments focused on specific transportation subsectors or particular geographic regions, for instance (Georgiadis et al., 2020; Fulzele and Shankar, 2023; Liu and Yuan, 2023; Chatziioannou et al., 2023; Zhang et al., 2023; Alper et al., 2015; Kutty et al., 2022), to name a few. In contrast, the evaluation of the performance of the transport sector at the European level is a topic that has been relatively underexplored in the literature. An overview of recent works on sustainability analysis at the European level is presented hereinafter.

A study by lo Storto and Evangelista (2023) analyzed the performance of land logistics in the EU, from 2010 to 2017, focusing on three main aspects: infrastructure efficiency, logistics quality, and environmental impact. The specific objectives include introducing a comprehensive benchmarking framework by using DEA, measuring the performance of EU countries in those areas. The study highlights that many countries struggle to improve both operational performance and environmental performance simultaneously.

Babaei et al. (2023) analyzed the sustainability performance of urban transportation systems across twelve European countries in a timeframe from 2001 to 2015, by using a decision support framework based on two DEA models, one evaluating countries annually and the second one assessing performance across all years simultaneously, and a multi-objective programming model to treat the models under uncertain conditions and also uncertainty in data. The efficiency scores are measured based on data related with factors including the levels of freight and passenger transportation, greenhouse gas emissions, energy consumption, and road accidents. Results showed that the best ranks varied based on uncertainties conditions.

Nikolaou and Dimitriou (2018) evaluated the road safety performance of 23 EU countries over a decade (2005–2014) by considering socio-economic and demographic factors that influence road traffic fatalities and serious injuries by using DEA and the DEA-Cross Efficiency Model (DEA-CEM). The study established both short-term and long-term road safety targets based on risk exposure indicators, providing a framework for policy improvements and strategic planning aimed at enhancing road safety across the EU. Also in this subject, a previous work by Shen et al. (2013) evaluated the road safety performance over a decade using the DEA and the Malmquist productivity index. The results of the work revealed that road safety progress in Europe was attained through the adoption of productivity-enhancing new technologies rather than through the relatively underperforming countries catching up with those best-performing ones.

A DEA-based framework was applied to assess the social sustainability for the regional EU road transport in 2004–2017 (Stefaniec et al., 2021) by using factors as mobility, accessibility, safety, employment, health, and equity. Authors demonstrated that under the social assessment perspective, new EU members perform better that the older ones, preventing bias due to inclusion of economic factors.

An overview of the EU member states and the United Kingdom transport sector was recently developed (Dolge et al., 2023) through a composite transport sustainability index, for cross-country comparison. Countries were compared using fifteen transport indicators grouped into four dimensions: mobility, sustainability, innovation, and environment. Additionally, a Logistic Mean Division Index (LMDI) decomposition analysis was also conducted to assess the changes in GHG emissions from the transport sector over a ten-year period. From the analysis outlining the good practice policies from the best performing countries and the combined results of the composite sustainability index and decomposition analysis, authors identified significant disparities between Nordic, Western, and Eastern European countries, emphasizing the need for tailored approaches to develop effective sustainability policies for the transport sector.

2.2. Sustainability measurement approaches

The sustainability performance can be measured through different approaches, which can encompass the traditional analysis of the multiple indicators, the structural analysis, the composite indicator calculation, the use of multi-criteria methods, and the use of DEA models.

The traditional analysis of the multiple indicators has been used by Liu and Yuan (2023), Eisenkopf and Burgdorf (2022). In the study described in Liu and Yuan (2023), eleven transport indicators were selected and constructed for the sustainable development goals (SDGs) under the United Nations indicator framework. The scores of each indicator were calculated, normalized and converted to scores of 0–100 and spatio-temporal patterns and interactions were analyzed. In Eisenkopf and Burgdorf (2022), various frequently discussed transport policy measures were examined with regard to their effects on the total transport performance of long-distance passenger transport within Germany and the transport performance of individual means of transport (modal split shares) in 2030. It used a simulation model based on system dynamics in which the transport performance for each simulated scenario is evaluated through multiple measures.

The method based on structural analysis (MICMAC) was used by Chatziioannou et al. (2023) to rank the key sustainable urban mobility indicators, enabling the identification of the most important indicators. Structural analysis is a tool used to identify and study system correlations to prioritize the pivotal variables or indicators for the functionality and evolution of the system. In other words, it enables the identification and systematic organization of all the key indicators via a matrix that presents their interrelations and clearly showcases their unique significance for the system.

A composite indicator (CI) is an increasingly prevalent method to synthesize masses of data for policy recommendation, evaluation and communication in a broad range of fields. The CI enables gathering complex and multi-faceted phenomena and, consequently, the comparison of the performance of several DMU (Cherchye et al., 2007). The CI is also an easy tool for communication purposes as it provides a big picture of a subject and, therefore, an easier interpretation. A well-known composite indicator is the Human Development Index (HDI), which ranks countries based on life expectancy, education and per capita income sub-indicators (Zhou et al., 2007).

The CI comprises several individual sub-indicators that measure different aspects of a phenomenon, which usually have no unit of measurement in common, summarizing them into a single index (Hoffmann and Giovannini, 2008). Eq. (1) shows the CI formulation in a general level, considering m sub-indicators i for each unit j, denoted by $y_{ij}(i = 1,\dots,m)$. When aggregating the sub-indicators i in a composite indicator, it is necessary to assign weights w_i to each of them according to the underlying theoretical framework, which enables the evaluation of the performance of unit j with respect to all m sub-indicators, through CI_j.

$$\mathrm{CI}_j = \sum_{i=1}^{m} w_i y_{ij} \tag{1}$$

The subjective judgement about the relative value of each sub-indicator is modeled through the weights w_i assigned to it Cherchye et al. (2007). Their values usually greatly impact the aggregation results from which the weighting method needs to be explicit and transparent (Joumard and Gudmundsson, 2010).

The easiest and most typical approach to attribute weights to the sub-indicators is the use of equal weights. This implies that all sub-indicators for all units have the same importance in the CI. However, this assumption is restrictive and typically not true. The assessment of the sustainability performance of the transport sector requires the comparison of multi-dimensional features of the transport sector in EU members. Thus, taking into account the uniqueness of each country, the adoption of a fixed set of weights is not considered a fair judgement, which is also remarked by Atkinson et al. (2002), as follows "in the context of the EU, there are evident difficulties in reaching agreement on such weights, given that each member state has its own national specificity".

Another practice to attribute weights is by incorporating experts or public opinions. Many methods can be used to compile and derive the weights from expert or public judgement, where a common approach is the use of multi-criteria methods, such as the Analytic Hierarchy Process (AHP) (Saaty, 1980). The AHP involves conducting pairwise comparisons between sub-indicators, during which experts are asked to assess pairs of sub-indicators and indicate which of the two is more important in their opinion, thus quantifying their relative importance. This approach, therefore, depends on the consistency of experts' opinions in the comparison matrix (Cherchye et al., 2007). By incorporating experts opinions, Fulzele and Shankar (2023) developed a performance index for a sustainable freight transportation system by innovatively integrating the Consensus Model (CM) with the Fuzzy Evidential Reasoning Algorithm (FERA). A CM was used to determine the degree of importance of each Key Performance Indicator identified across three dimensions of sustainability, owning the capability to appropriately manage the non-cooperative behaviors and minority opinions of the experts. FERA has been used to aggregate subjective judgements provided by experts with crisp quantitative values. This approach has a unique capability to handle various uncertainties related to impreciseness in decision-making.

Finally, an alternative method to determine weights of sub-indicators is to derive them endogenously through the Data Envelopment Analysis models (Zhou et al., 2007) in order to avoid the subjectivity related to weight choice, being possible to reflect the national specificity of EU members. The DEA is a non-parametric method, proposed by Charnes et al. (1978), that evaluates the relative efficiency of several decision-making units (DMU) based on linear programming. It measures the efficiency of each DMU, given observations on inputs and outputs in a set of homogeneous entities, without knowledge of the production or cost function (Cherchye et al., 2007).

The DEA promotes measuring the DMU efficiency in terms of the Pareto-Koopmans concept. Thus, a unit is efficient when any increase in any output cannot be obtained without decreasing, at least, another output or increasing, at least, another input. Alternatively, a unit is efficient when a decrease in any input cannot be obtained without increasing at least another input or decreasing at least another output (Fried et al., 1993).

By comparison with the best practices frontier, the DEA model promotes the weights' selection that is the most advantageous for DMU under assessment (Reisi et al., 2014). This means that the weights are derived from the data itself, avoiding *a priori* assumptions and computations involved in fixed weight choices (Cooper et al., 2007). Accordingly, DEA is a popular method in the CI literature as it can solve the problem of subjectivity in the weighting procedure.

Another well-known property of the DEA model is its unit invariance. This property constitutes an advantage for the construction of the CI as its final value is independent of the measurement units of the sub-indicators, which in turn makes the normalization stage redundant and unnecessary (Cherchye et al., 2008).

As a by-product of DEA, each DMU is classified as efficient or inefficient. The linear combination of the efficient DMUs creates the best practices frontier used as a reference to calculate the efficiency of each unit based on its distance to the frontier. The efficient DMUs on the frontier have an efficiency score equal to 1 (or 100%), while the DMUs outside the frontier, the inefficient ones, have an efficiency score less than 1. The envelopment part of the DEA comes from this property since the frontier is said to envelop the observed DMUs (Cooper et al., 2007).

In the DEA approach, the models require the definition of the multiple inputs and outputs for a set of homogeneous units to evaluate their efficiency. In the context of sustainability performance, some examples are Georgiadis et al. (2020), Kutty et al. (2022), Zhang et al. (2023) and Gruetzmacher et al. (2021b).

Georgiadis et al. (2020) evaluates the performance of 34 multimodal public transport networks involving the bus and metro systems. A model with constant returns to scale is defined for each system, which involves the definition of inputs and outputs and output orientation. Kutty et al. (2022) proposes a Slack Based Measure DEA model to handle undesirable factors both in the inputs and outputs in the technology set. This model evaluates the integrated relative sustainability performance assessment to determine the most efficient smart city under the six dimensions of sustainable development. Zhang et al. (2023) develops two sets of methods that consider only the energy factor and multiple other factors, including energy, to evaluate and analyze the carbon reduction performance of China's provincial transportation sector. This paper exploits the DDF model by including a set of input and output indicators to evaluate the transportation carbon efficiency in various regions.

Gruetzmacher et al. (2021b) evaluates the environmental performance of the transport sector in 28 European Union countries from 2015 to 2018 through the BoD model proposed by Färe et al. (2019), in which a composite indicator is calculated by aggregating several sub-indicators, without the need to identify them as inputs and outputs.

3. Methodology

This section describes two alternative Benefit-of-the-Doubt (BoD) models used to derive composite indicators by aggregating a set of isotonic and anti-isotonic sub-indicators (Dyson et al., 2001), capturing positive and negative aspects of performance, respectively.

The initial DEA model to compute the CI is referred to as the BoD model, which was proposed by Cherchye et al. (2007). This model handles only a set of isotonic sub-indicators, requiring the transformation of anti-isotonic sub-indicators to capture overall aspects of the performance. Alternatively, the model proposed by Färe et al. (2019) enables to handle directly both isotonic and anti-isotonic sub-indicators. Hereafter, this model is shortened by the initials of the authors' names, Färe, Karagiannis, Hasannasab and Margaritis, as the FKHM model.

These two models are applied to compute the composite indicator (CI) and examine their differences in performance assessment through a small illustrative example. Additionally, the example explores various strategies for transforming anti-isotonic data into isotonic form.

3.1. Benefit-of-the-doubt model

The BoD model (Cherchye et al., 2007) is equivalent to the original DEA input-oriented model, with all sub-indicators considered as outputs and a single dummy input equal to one for all units. The dummy input can be understood intuitively by regarding the model as a tool for aggregating several sub-indicators of performance, without referencing the inputs that are used to obtain this performance (Chung, 2017). Since the BoD model only includes outputs, it measures the DMU performance rather than its efficiency.

3.1.1. Formulation

Given a cross-section of m isotonic sub-indicators and s DMUs, with y_{ij} being the value of sub-indicator i for the DMU j, and w_i the weight attributed to the ith sub-indicator. These weights are endogenously calculated using the BoD model (2) for each DMU under evaluation j_0, enabling to compute its composite indicator (CI_{j_o}) to evaluate its performance through the weighted average of m sub-indicators, as follows:

$$
\begin{aligned}
&\mathrm{CI}_{j_0} = max \sum_{i=1}^{m} w_i y_{ij_0} \\
&\textbf{\textit{s.t.}} \qquad \sum_{i=1}^{m} w_i y_{ij} \le 1 \qquad \forall j = 1, \ldots, s \\
&\qquad\qquad w_i \ge 0 \qquad \forall i = 1, \ldots, m
\end{aligned}
\tag{2}
$$

For each DMU j_0, the model chooses the non-negative w_i for each sub-indicator y_{ij_0} that maximizes its CI_{j_0} score by comparison with the best practices frontier. The core idea supports that if a sub-indicator has a good (poor) relative performance for a specified unit, this dimension should be emphasized (curtailed) and the model (2) tends to assign a higher (lower) weight to it. Therefore, a DMU cannot claim that a defective relative performance is due to a harmful or unfair weighting scheme, since the model (2) maximizes the CI for each DMU under assessment by comparison with best practices frontier, allowing the weights w_i to be endogenously estimated (Cherchye et al., 2007).

In some situations, however, a DMU may obtain a higher relative performance by assigning zero weights to sub-indicators with lower scores. In this way, the sub-indicators associated with the zero weights would not influence global performance, and the CI scores might end up being based on a small subset of all the sub-indicators (Dyson et al., 2001). To prevent sub-indicators from being under-emphasized, the model should incorporate additional restrictions for the sub-indicator weights. Then, virtual proportional weight restrictions (Wong and Beasley, 1990) (*i.e.* sub-indicator share restrictions) given by (3) are imposed in the model (2) to assess the sustainability performance of the transport sector, following Färe et al. (2019), in which a lower bound to the sub-indicator share is defined while keeping the upper bound free.

$$
\frac{w_i y_{ij_0}}{\sum_{i=1}^{m} w_i y_{ij_0}} \ge \alpha \ \forall i = 1, \ldots, m \tag{3}
$$

Thus, each sub-indicator is required to have a minimum contribution percentage α in the assessed composite indicator for the country under evaluation.

3.1.2. Transformation methods

As previously mentioned, the BoD model (2) derives the composite indicator for each DMU under assessment by aggregating isotonic sub-indicators. However, it is frequent that the performance assessment has to handle anti-isotonic sub-indicators, which need to be transformed before being incorporated in the BoD model and handled as the isotonic sub-indicators (Färe et al., 2019).

Some of the most common data transformation methods include the inverse of the value of the anti-isotonic sub-indicator (inversion method) (Lovell et al., 1995), the subtraction of the anti-isotonic sub-indicator from a sufficiently positive constant, K (constant method) and the rescaling normalization of the anti-isotonic sub-indicator using the maximum–minimum (MM) method. These methods are presented and compared in Dyson et al. (2001) and Scheel (2001), and further discussed in the context of the BoD model (2), in the Section 3.3.

The inversion method, however, annihilates the ratio or interval scale of the data (Dyson et al., 2001), while the constant method is sensitive to the choice of the constant K, as a large value can dominate the data and change the best practices frontier (Dyson et al., 2001). Then, a smaller value K is more adequate as it reduces the effect of the translation on the results (Dyson et al., 2001). In the maximum–minimum (MM) method, the anti-isotonic sub-indicator is normalized through the ratio between the difference of each anti-isotonic sub-indicator from its sample maximum score over the range value observed on the sample.

Since the BoD model is derived from an input-oriented DEA model with constant returns to scale, it is not translation invariant for the sub-indicators values. This implies that the use of translated or rescaled data can identify different best practice frontiers, affecting the CI results and, consequently, the ranking of the DMUs (Färe et al., 2019), which can imply misleading interpretations concerning their true performance.

3.2. FKHM model

Färe et al. (2019) modified the original BoD model to directly incorporate the anti-isotonic sub-indicators, obtaining the FKHM model, avoiding any data transformation. This model treats the anti-isotonic sub-indicators as reverse rather than as undesirable. This means that the model assumes that the reverse sub-indicators values can vary independently from the values of isotonic sub-indicators considered forward sub-indicators. As argued by Färe et al. (2019), the presence of forward indicators does not imply nor is implied by the presence of reverse indicators. Given a cross-section of M sub-indicators and s units, y_{ij} is the value of sub-indicator i for the unit j, and w_i is the weight attributed to the sub-indicator i. The formulation for the FKHM model is presented in (4), where y_{ij} $(j = 1, \dots, m)$ are the forward sub-indicators (*i.e.*, capturing positive aspects in performance) and y_{ij} $(j = m+1, \dots, M)$ are the reverse sub-indicators (*i.e.*, capturing negative aspects in performance).

$$\begin{aligned} \mathrm{CI}_{j_0} &= max \sum_{i=1}^{m} w_i y_{ij_0} - \sum_{i=m+1}^{M} w_i y_{ij_0} \\ \textit{s.t.} \quad & \sum_{i=1}^{m} w_i y_{ij} - \sum_{i=m+1}^{M} w_i y_{ij} \le 1 \qquad \forall j = 1, \dots, s \\ & w_i \ge 0 \qquad \forall i = 1, \dots, M \\ & y_{ij} \ge 0 \qquad \forall i = 1, \dots, M, \ \forall j = 1, \dots, s \end{aligned} \tag{4}$$

The main difference between the BoD model (2) and the FKHM model (4) is that the first one maximizes the weighted average of forward (isotonic) sub-indicators, while the second maximizes the difference between the weighted average of forward sub-indicators and the weighted average of reverse sub-indicators of the DMU under assessment. Additionally, the presence of forward sub-indicators does not imply the presence of reverse ones. Thus, when there are no anti-isotonic indicators, the FKHM model can be reduced to the formulation of the BoD model (Färe et al., 2019).

Additionally, it is necessary to prevent sub-indicators from being under-emphasized in the evaluation of the sustainability performance of the transport sector in European countries by incorporating the virtual proportional weight restrictions (5), in the model (4) following (Färe et al., 2019).

$$\frac{w_i y_{ij_o}}{\sum_{i=1}^{m} w_i y_{ij_0} + \sum_{i=m+1}^{M} w_i y_{ij_0}} \ge \alpha, \ \forall i = 1, \dots, M \tag{5}$$

Table 1
Dataset of the example.

Unit	y_1	y_2
A	14.4 (6th)	102.7 (8th)
B	33.7 (1st)	103.1 (9th)
C	14 (7th)	67 (6th)
D	32.9 (2nd)	71.4 (7th)
E	29.03 (4th)	63 (4th)
F	15 (5th)	40 (3rd)
G	4.61 (8th)	10 (2nd)
H	33.7 (1st)	110 (10th)
I	31 (3rd)	65 (5th)
J	4.61 (8th)	5 (1st)

Table 2
FKHM and BoD results of the example.

	Model			
	BoD-inversion	BoD-MM	BoD-constant	FKHM
A	0.452 (8th)	0.435 (6th)	0.435 (6th)	0.427 (4th)
B	1 (1st)	1 (1st)	1 (1st)	1 (1st)
C	0.463 (7th)	0.679 (5th)	0.679 (5th)	0.415 (5th)
D	0.996 (2nd)	1 (1st)	1 (1st)	1 (1st)
E	0.894 (4th)	0.980 (2nd)	0.980 (2nd)	0.882 (2nd)
F	0.535 (6th)	0.910 (4th)	0.910 (4th)	0.450 (3rd)
G	0.566 (5th)	0.960 (3rd)	0.960 (3rd)	0.140 (6th)
H	1 (1st)	1 (1st)	1 (1st)	1 (1st)
I	0.948 (3rd)	1 (1st)	1 (1st)	1 (1st)
J	1 (1st)	1 (1st)	1 (1st)	1 (1st)
Mean	0.785	0.896	0.896	0.732

3.3. Illustrative example

A comparison analysis of the different transformation techniques to deal with anti-isotonic, considered as reverse sub-indicator is addressed by using a simplified, hypothetical, example comprising 10 DMUs (A to J) with a single forward sub-indicator (y_1) and a single reverse sub-indicator (y_2). Table 1 presents the data used in the example, and the ranking for each unit concerning each sub-indicator is indicated in brackets.

The CI was calculated using the BoD model (2) after transforming the anti-isotonic sub-indicator y_2 using the inversion, constant and MM methods. In this example, to avoid the transformation of any sub-indicator to obtain a zero value in the BoD-MM method, the maximum score of that sub-indicator was replaced by a constant, K, 10% higher than the sample maximum. Since this score is suitable in the BoD-constant method, it is chosen to transform the y_2 into an isotonic sub-indicator.

Table 2 shows the CI results obtained for each unit (A to J) when applying the BoD model with different transformation methods, *i.e.*, BoD-inversion, BoD-MM and BoD-constant, and the FKHM model. Additionally, the ranking for each unit is included in brackets concerning its performance for each model.

The results obtained with the different transformation methods are graphically presented. Fig. 1(a) shows the BoD model results with the inversion of the sub-indicator y_2. Fig. 1(b) shows the BoD model results with the subtraction of sub-indicator y_2 from constant K, which value was set as K = 121. Fig. 1(c) shows the CI results using the maximum–minimum (MM) method, setting the maximum value as $K = 121$. Finally, Fig. 1(d) presents the results of the FKHM model that includes y_2 as a reverse sub-indicator.

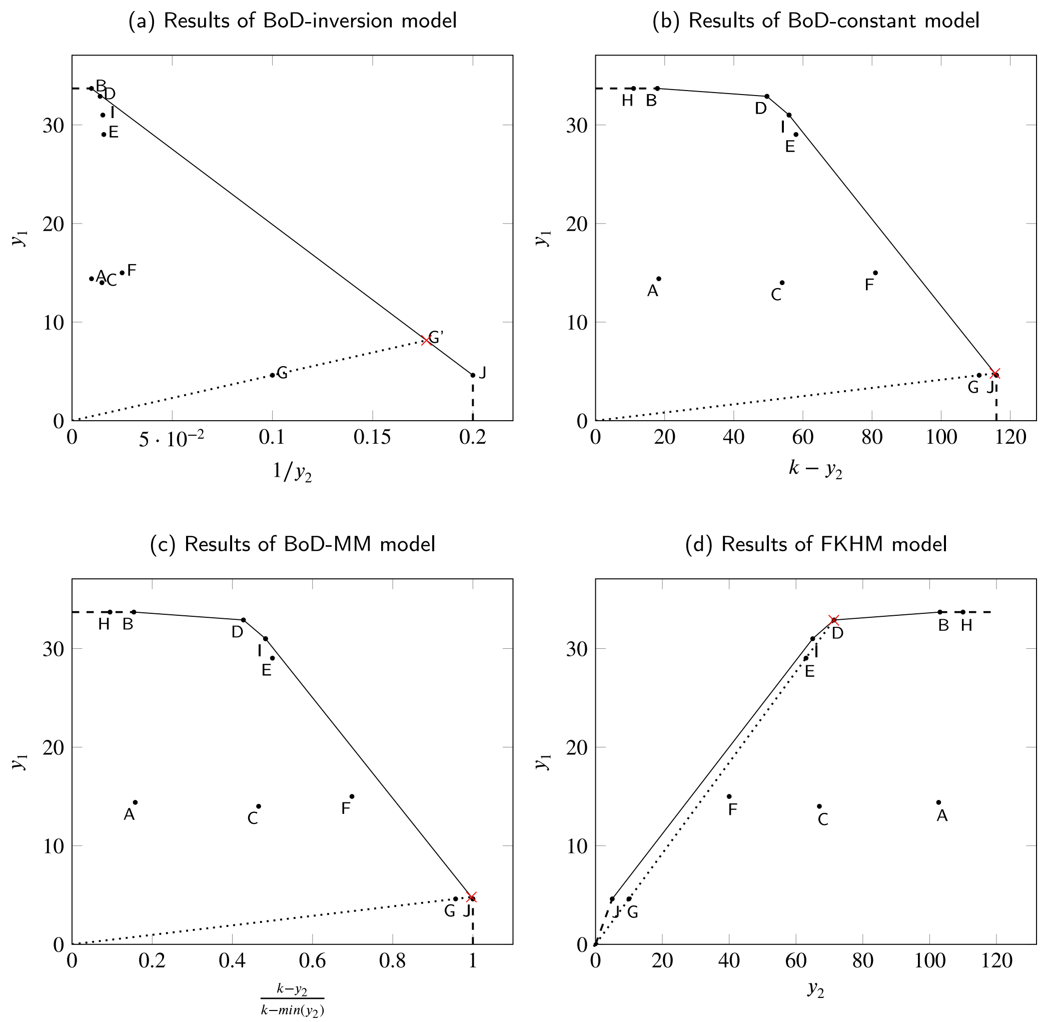

Fig. 1. Results of BoD-inversion (a), BoD-constant (b), BoD-MM (c) and FKHM (d) models.

In this example, the BoD-MM and the BoD-constant methods allow to obtain the same CI values, since they employ equivalent approaches in the transformation of the anti-isotonic sub-indicator, by subtracting it from a constant, because the K used in the BoD-constant method is equal to the y_{max} considered in the BoD-MM method. In fact, the normalization step, seen in the BoD-MM method, is redundant in DEA models and does not change the CI results (Cherchye et al., 2008). Hereinafter, the results achieved on BoD constant, BoD inversion and FKHM models are analyzed.

These approaches to deal with the anti-isotonic sub-indicator generate different CI results, although the Kruskal–Wallis test indicates that there are no significant differences between them ($p_{value} = 0.5711$).

Units B, D, I and J are Pareto-Koopmans efficient in BoD-constant and FKHM models, except for units D and I in the BoD-inversion model, which have $CI = 0.996$ and $CI = 0.948$, respectively. In fact, it is observed that BoD constant and FKHM models have the same benchmarks which define the best practice frontier, while the BoD inversion model identifies only B and J as benchmarks, although D and I are assessed with high-efficiency scores. This result confirms that inversion transformation destroys the ratio or interval scale of the data (Dyson et al., 2001) and it should be avoided.

It is observed that unit B has the best ranking in the y_1, and it is the 9th ranking in y_2, being efficient in all models. This occurs since unit B has the best ranking regarding y_1, the forward sub-indicator, and, by default, it is compared with itself since it has the lowest score of the anti-isotonic sub-indicator for this level of the forward sub-indicator. The same occurs for the efficient unit J, which has the lowest reverse sub-indicator for the observed level of the forward sub-indicator, being compared with itself. It should be noted that unit H is inefficient in all models, as it is compared with B, owning a slack associated with the anti-isotonic sub-indicator.

In the example under analysis, the highest difference between the results of the models presented in Table 2 relates to unit G. Comparing with unit J, that unit has worse performance in the sub-indicator y_2, having the same y_1, denoting that it should be evaluated as inefficient. This result is reflected in the FKHM model by assessing G with the lowest efficiency (0.140), while the BoD-constant model evaluates it with higher efficiency (0.96), and the BoD-inversion model assesses it with lower efficiency than the average (0.566). Although it is straightforward to understand the efficiency scores obtained in the BoD models, projecting G against the frontier to achieve the efficient target (plotted by red x in each figure), it is interesting to detail the behavior of the FKHM model.

For each unit (y_1, y_2) evaluated in the production possibility set, the FKHM model finds the units that show the highest slope y_1/y_2 to define the best-practice frontier. In this example, these units are J, I, D and B, which define the linear segments of the best practice frontier. Therefore, each inefficient unit's observed slope y_1/y_2 is projected against the frontier. In the example, the unit G of Fig. 5 is projected on the red x point given by (71.31, 32.87), being compared with the benchmarks I and D, obtaining the $CI = \frac{OG}{OG'} = 0.140$. Thus, the efficient target indicates an increased level of both sub-indicators for the efficient levels observed on the frontier. To achieve the observed slope on the inefficient unit, the distance from it to achieve the frontier tends to be higher when using FKHM model than when using BoD-inversion and BoD-constant approaches. This explains why the lowest CI is observed for unit G in the FKHM model. This effect tends to increase the average distance against the frontier observed on inefficient units in the FKHM model.

Although there are slight differences observed in the CI results, it also is verified that the best practice frontier is defined by the same efficient units in the FKHM and the BoD-constant models in this example. This increases the robustness of these models to the decision makers' point of view in using them to assess the performance when multiple forward and reverse sub-indicators are taken into account. In the multiple forward and reverse scenario, it is possible to infer that the FKHM model assesses with high efficiency scores the units which have higher levels of forward indicators and, at the same time, lower levels of reverse sub-indicators. Additionally, inefficient units that have lower levels of forward and reverse sub-indicators such as G unit, tend to have a lower score of efficiency since the efficient units of the frontier dominate them. Supported by these comparison analysis results, the methodological approach decision is made by considering the FKHM and the BoD-constant models to assess the sustainability performance of the transport sector in EU members. The BoD-Constant approach was hereafter named as BoD(K).

4. Case study

The performance assessment of the transport sector in EU countries using the CI, derived through the FKHM and BoD(K) models, is conducted for the period between 2015 and 2018. This time span precedes the disruptive impacts of the Covid-19 pandemic, ensuring unbiased data. The selection of the sub-indicators plays a critical role in computing the overall performance of the transport sector in EU countries. This procedure was driven to provide a global analysis of the transport sector, *i.e.*, non-segmented nor non-modal analysis, while encompassing the key dimensions of the problem, and ensuring data availability for all units under analysis. Previous research (Gruetzmacher, 2021; Gruetzmacher et al., 2021a) grounded in similar conceptual frameworks, *i.e.*, the transport sustainability goals outlined in the 2011 White Paper (European Commission, 2011) and the SDGs from the United Nations (United Nations, 2022), assisted in defining the sub-indicators, as outlined below.

4.1. Data and variables

The proposed CI aims to achieve a balance between what is necessary to support a sustainable transport assessment and the available data for EU countries. Besides, each selected sub-indicator must be easy to interpret and should measure a specific dimension of the performance, ensuring a minimal number of sub-indicators while guaranteeing they convey the most important dimensions of the problem. All the data used in this work were gathered from the Eurostat database (Eurostat, 2021).

The CI for assessing the transport sustainable performance of European countries, belonging to the EU, was constructed based on three forward sub-indicators (*i.e.*, capturing positive aspects) and four reverse sub-indicators (*i.e.*, capturing negative aspects).

The forward sub-indicators are the share of buses and trains in total passengers' transport (*public transport*), the share of energy from renewable sources in transport (*renewable fuels*) and the share of rail and inland waterways in total freight transport (*freight transport*). The reverse sub-indicators include the people dead in road accidents (*road deaths*), the GHG emissions by fuel combustion in transport (*GHG emissions*), the average carbon dioxide (CO_2) emissions per kilometre from new passengers cars (*new car emissions*) and the energy dependency on oil and petroleum products (*energy dependency*). These sub-indicators are described hereinafter.

Public transport measures the share of passenger's transport made by buses (including coaches and trolleybuses) and trains in the total inland transport (buses, trains and passenger cars), being expressed in percentage. Trams and metros are not included due to the lack of harmonized data. This sub-indicator relates to the necessity of improving transport quality, accessibility and reliability, as discussed in the White Paper on Transport: Roadmap to a Single European Transport Area (European Commission, 2011) and the Sustainable Development Goals (United Nations, 2022). Higher shares of public transport relate with building resilient and sustainable infrastructure and renewing and planning cities to provide access to basic services for all.

Renewable fuels are expressed as the percentage of renewable fuels in the total transport fuels. Energy by renewable sources consumed in transport is given by the sum of sustainable biofuels, renewable electricity, hydrogen and synthetic fuels of renewable origin, and other reported forms of renewable energy (European Commission, 2018b). The White Paper on Transport (European Commission, 2011) suggests a regular phase-out of conventionally-fueled vehicles from urban environments by halving their number in 2030 and phasing them out of the cities by 2050, and the Renewable Energy Directive promotes an increase to 32% of renewable energy in the transport sector for 2030 (European Commission, 2018a). The *renewable fuels* sub-indicator shows how expansive is the use of renewable energy in the transport sector and, therefore, aids in assessing the performance of the countries towards these targets.

Freight transport is expressed in percentage and measures the share of freight on the national territory of these modes in the total inland transport (road, rail and inland waterways). Sea and air freight transport are included. The *freight transport* sub-indicator was not applicable to Cyprus and Malta since these countries did not present values for railways or inland waterways. As an effort to have a complete database without excluding these countries from the evaluation in this work, the lowest values observed on the dataset were used for Cyprus and Malta in the years under analysis. This approach prevents those countries from becoming unintended benchmarks, and it has also been used by Morais and Camanho (2011). This sub-indicator is closely related to the Roadmap target of shifting 30% of the road freight to other modes, such as rail and waterways, by 2030 and more than 50% by 2050.

Road deaths measures the number of fatalities in road accidents per hundred thousand inhabitants. This sub-indicator includes passengers, drivers of motorized vehicles, pedal cycles, and pedestrians, who have died up to 30 days after the accident. This sub-indicator is aligned with the Sustainable Development Goals aim of safer and healthier cities, with well-being status. The Roadmap also highlights the EU's aim to reduce road fatalities close to zero by 2050.

GHG emissions measures the transport's fuel combustion contribution in the total greenhouse gas emissions inventory. The values, in thousand tonnes, were normalized using the countries' population on 1st January of each year, to take into consideration their dimension. Since the 2011 White Paper on transport (European Commission, 2011) sets out a target of 60% reduction in GHG emissions by 2050 compared to 1990 levels, it is important to be able to assess the countries' performance towards this target.

New car emissions is defined as the average weight, in grams, of CO_2 per kilometre in a given year for new passenger cars. This is a target for the average of the manufacturer's overall fleet, meaning that cars above the limit are allowed in the market as long as the production of lighter cars offsets them. The Regulation (EU) 2019/631 sets a mandatory target for emission reduction for new cars of 95 g of CO_2/km by 2021 (European Commission, 2019a).

Energy dependency monitors to which extent the countries' economies rely on oil and petroleum products imports to meet their energy needs. It is calculated by dividing the net imports by the gross available energy. It is used on a percentage basis, from which an energy dependency value may be higher than 100%, for countries creating a stock in a given year, or negative, for oil exporter countries.

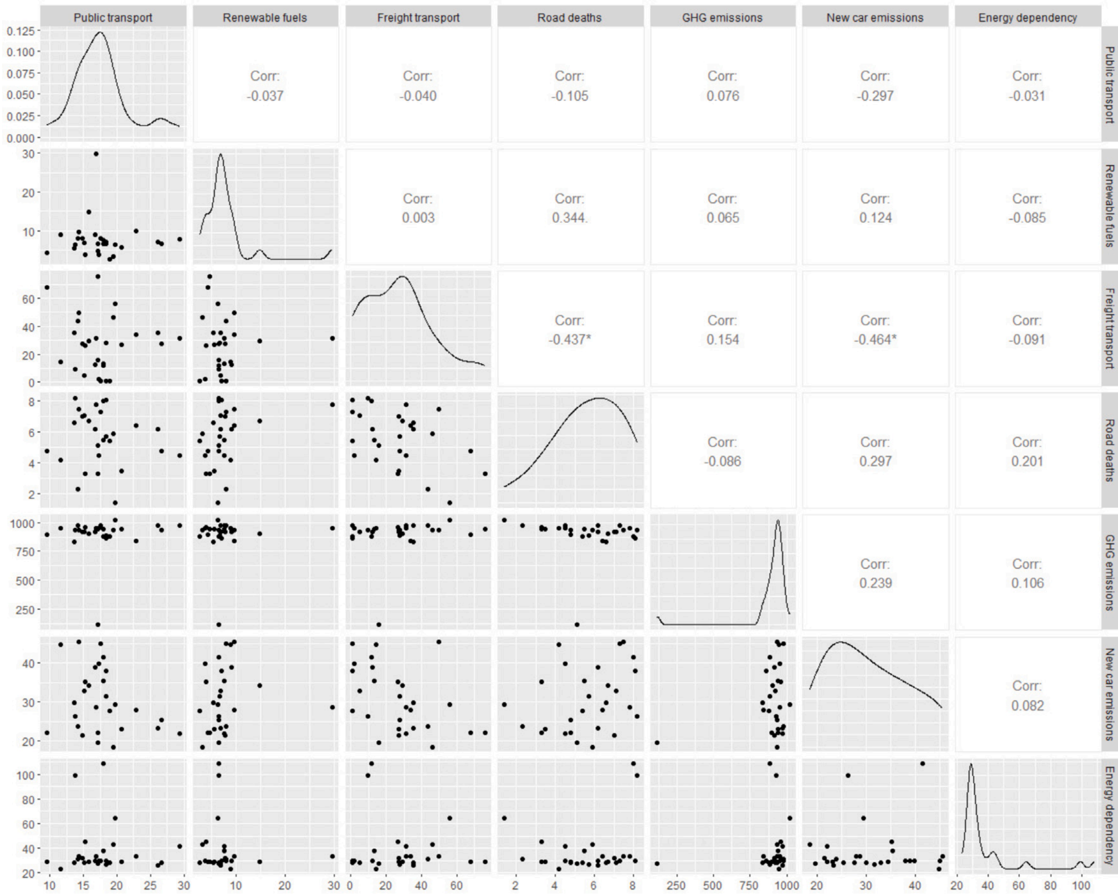

Fig. 2. Correlation of the sub-indicators in 2018 (Significance level: '.' 10%, '*' 5%, '**' 1%).

A negative value occurred only once in the dataset for an exporter country (−4.701%). To maintain the relative position of all units regarding the other countries, a constant K of 5.701% was added to each energy dependency sub-indicator score. Thus, the best score of 1% is assigned to the exporter country, keeping the same relative position of the units. As oil becomes scarcer each year, implementing resource-efficient policies without reducing the transport system efficiency, is fundamental and one of the objectives mentioned in the Roadmap.

These seven sub-indicators are used to assess the transport performance of EU countries, as presented in the next section.

4.2. Descriptive analysis of the variables

The performance of the transport sector was assessed for the 28 EU countries, from 2015 to 2018. This assessment was conducted, taking into account data from all 112 DMUs. The study considers the United Kingdom data since it still integrated the European Union in the time span of the assessment. Table 3 shows the mean of the sub-indicators, the dispersion coefficient (DC) and the mean variation between 2015 and 2018 in percentage.

Table 3
Mean, DC and variation of the sub-indicators values.

Sub-indicator	Mean	DC	Variation
Public transport	17.910	0.242	+2.126%
Renewable fuels	6.844	0.726	+17.238%
Freight transport	27.230	0.727	−3.257%
Road deaths	5.537	0.348	−7.376%
GHG emissions	213.239	0.724	+3.820%
New car emissions	119.832	0.069	−0.405%
Energy dependency	96.475	0.230	−3.028%

Analyzing the forward sub-indicators in Table 3, it can be seen that the average of the *Public transport* and the *Renewable fuels* sub-indicators have increased between 2015 and 2018. The highest improvement was observed on the *Renewable fuels* sub-indicator, increasing more than 17%. The *Freight transport* sub-indicator was the only forward sub-indicator that had a decrease in the time span under study, being −3.257% on average. Regarding the reverse sub-indicators, the *Road deaths*, *New car emissions*, and the *Energy dependency* sub-indicators had a decrease in their average in 2018 when compared to 2015. However, the average of *GHG emissions* has increased and, by 2018, was almost 4% higher than 2015 levels. The sub-indicators *Renewable fuels*, *Freight transport* and *GHG emissions* presented the highest values of DC, or data dispersion relative to the mean. The high DC in the *Renewable fuels* sub-indicator can be attributed to the difference among countries in exploring the available renewable energy. The high DC in *Freight transport* can be related to some countries' advantageous geographical locations and environmental conditions that facilitate the utilization of rail and inland waterways. Additionally, the high DC for the *GHG emissions* sub-indicator might reflect the different policies adopted by the EU countries for reducing emissions. The lowest DC was observed for the *New car emissions* sub-indicator, which could capture the higher homogeneity for manufacturers operating in the EU market to accomplish the mandatory target for emission reduction.

Fig. 2 presents the correlation scores for each pair of sub-indicators in 2018. The absolute values of the correlations exhibit low magnitudes, ranging from 0.003 to 0.464, with only three statistically significant scores. These include *Road deaths* and *Renewable fuels* (0.344, $p.value = 0.073$), *Road deaths* and *Freight transport* (−0.437, $p.value = 0.020$), and *New car emissions* and *Freight transport* (−0.464, $p.value = 0.013$). A

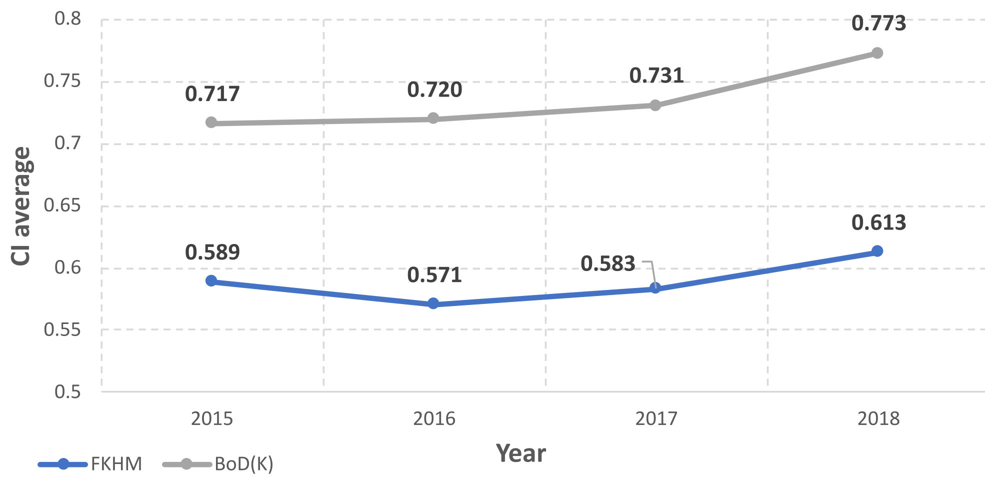

Fig. 3. Evolution of the average of the EU countries' CI scores from 2015 to 2018.

similar pattern of correlations is observed in previous years. Despite the low magnitudes of the correlations, all sub-indicators were retained in the model since they do not significantly correlate.

5. Performance assessment of the transport sector in EU

5.1. Results

The CI to assess the transport sector performance for each country and year was calculated using the data from the time span of four years, which results from the comparison with the best practices observed during this complete time period. Thus, a country evaluated in a given year is considered as a different DMU, implying that each model includes 112 DMUs corresponding to the 28 countries multiplied by the number of years. Given that each model includes 112 DMUs, seven indicators (outputs), and a single input (equal to 1), it adheres to the rule of thumb proposed by Dyson et al. (2001), which states that the number of DMUs should be at least twice the total number of inputs and outputs.

It was defined that the sub-indicator share restrictions (3) are imposed in BoD(K) model (2). Similarly, the sub-indicator share restrictions (5) are imposed in the FKHM model (4). These sub-indicator share restrictions should have a contribution of at least 5%, $\alpha = 0.05$. This value was chosen to prevent the attribution of zero weights at any sub-indicator, guaranteeing all sub-indicators contribution in the final CI. Moreover, a higher weight, such as $\alpha = 0.10$, was tested; however, it was not employed as this value tends to penalize countries underperforming in a single sub-indicator.

The averages of the CI for both models, BoD and FKHM, in the time span under analysis, are shown in Fig. 3.

When using the BoD(K) model, ten units were considered efficient: Denmark (2015 and 2017), Latvia (2018), Hungary (2015, 2016 and 2017), Netherlands (2015 and 2018) and Sweden (2017 and 2018). The average of the CI scores in the four years when using this model was 0.735. The CI varied slightly during these four years, settling in 2018 at 0.773, 7.80% above 2015 levels. The CIs obtained with the BoD(K) model for each country and year are presented in Fig. 4.

With the FKHM model, twelve units were efficient: Denmark (2017), Latvia (2015 and 2018), Hungary (2015, 2016 and 2017), Netherlands (2015, 2017 and 2018), Romania (2015) and Sweden (2017 and 2018). The average of the CI results in the four years analyzed was around 0.589. Through the FKHM model, the average had decreased by almost 3% in 2016 when compared to 2015 but began to increase in 2017. By 2018, the average of the CI was 4.09% above 2015 levels. Fig. 5 shows the CI results obtained with the FKHM model for each country and year analyzed.

Table 4 summarizes the benchmarks and the number of inefficient units that should emulate their best practices, categorized by sub-indicator, for both models. Based on the results of the BoD(K) model, Sweden (2017 and 2018) and Denmark (2017) emerge as benchmarks of 32, 60 and 38 inefficient units, respectively. Sweden (2018) had the highest level of *Renewable fuels* and a low level of *Road deaths*. Sweden (2017) achieved the minimum *Road deaths*, the second highest level of *Renewable fuels*, and ranked in the best percentile (BP) for *Energy dependency*. Considering the results of the FKHM model, Hungary (2015), Latvia (2015) and Sweden (2018) are the benchmarks of 67, 55 and 84 inefficient units, respectively. Sweden (2018) stands out for its high level of *Renewable fuels* and a low level of *Road deaths*. Hungary (2015) achieved the maximum performance in *Public Transport* while maintaining low *GHG emissions* (BP). Similarly, Latvia (2015) demonstrated strong performance in *Freight transport* (BP) alongside low level of *GHG emissions* (BP).

Comparing the average of CIs derived from both models over time, their patterns are similar. However, the CI average derived from the FKHM model is almost 20% lower than the average CI obtained with the BoD(K) model. Nine units can be considered robust benchmarks since they are efficient in both models: Denmark (2017), Latvia (2018), Hungary (2015, 2016 and 2017), Netherlands (2015 and 2018) and Sweden (2017 and 2018).

The most noticeable differences between the BoD(K) and the FKHM results were in the CI results for Portugal and the United Kingdom, in which the FKHM results were more than 50% lower than the BoD(K) results. There were also eleven other countries in which the CI scores were more than 30% lower in the FKHM model when compared to the BoD(K) model: Belgium, Germany, Estonia, Ireland, Greece, Spain, France, Croatia, Italy, Cyprus and Malta. All these countries had better performances on reverse sub-indicators and not so remarkable values on the forward sub-indicators. In fact, those countries have at least one forward sub-indicator that is lower than its average. These results corroborate the previous idea that the FKHM model tends to decrease the CI score of the inefficient units that have lower scores of forward and reverse sub-indicators.

The analysis of the areas in which the inefficient units were outperformed by the efficient ones (benchmarks) is performed through a radar analysis, from which it is possible to reveal possible areas of improvement for the first ones. Therefore, the radar analysis facilitates the comparison of the average for each sub-indicator between benchmarks and inefficient units, incorporating data from all four years under consideration.

Fig. 6 compares the performance of the benchmarks obtained using the BoD(K) model. The forward sub-indicators are highlighted in green, while the anti-isotonic sub-indicators are highlighted in red.

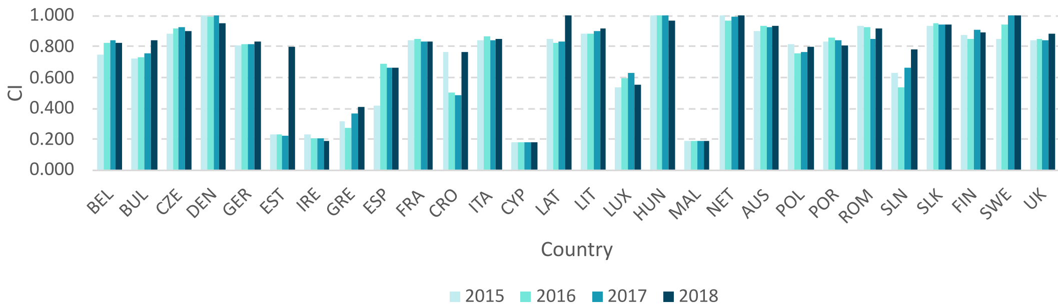

Fig. 4. CI scores from 2015 to 2018 obtained with the BoD(K) model.

Table 4
Benchmarks and inefficient units in BoD(K) and FKHM models.

Benchmarks	BoD(K)	FKHM	*Public transport*	*Freight transport*	*Renewable fuels*	*GHG emissions*	*Energy Dependency*	*Road deaths*	*New car emissions*
HUN15	1	67	1st high			BP			
HUN16	14	0	2nd high			BP	BP		
HUN17	15	1	3rd high	BP	BP	BP	BP		
LAT15	–	55		High		BP			
LAT18	27	7		3rd high					
ROM15	–	3	BP	BP		1st low	BP		
DEN15	4	–					BP	BP	BP
DEN17	38	4					1st low	2nd low	BP
SWE17	32	3			2nd high		BP	1st low	
SWE18	60	84			1st high			low level	
NET15	6	0		BP				BP	BP
NET17	–	1		BP			BP	BP	BP
NET18	18	0		BP	BP			BP	BP

BP: best percentile.

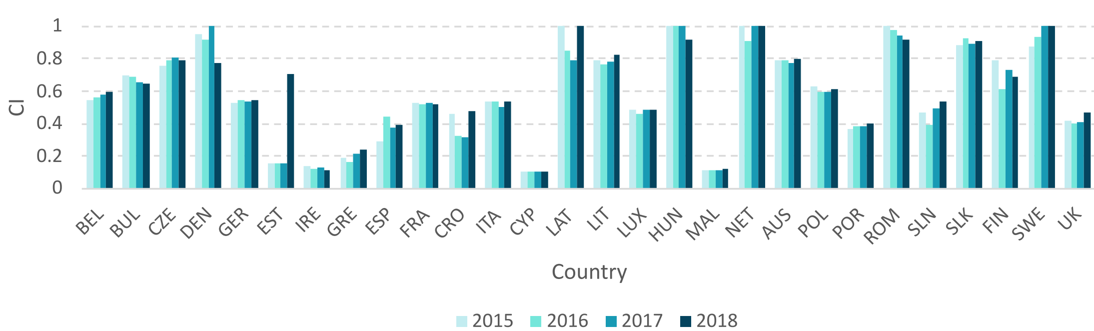

Fig. 5. CI scores from 2015 to 2018 obtained with the FKHM model.

Analyzing the results in Fig. 6 the inefficient units were outperformed by the benchmarks in all sub-indicators except for the *new car emissions*, in which they present similar levels. For the forward sub-indicators, the inefficient units have less than 60% of the share of *renewable energy* achieved by the benchmarks, only about 70% of the share of *freight transport* and 84% of the share of *public transport*. For the anti-isotonic sub-indicators, in red, the inefficient units had values 25% higher in *road deaths*, 27% in *GHG emissions* and 22% in *energy dependency* than the benchmarks units.

Fig. 7 confers a similar analysis by comparing the benchmarks and the inefficient units obtained from the FKHM model.

It is possible to conclude that the inefficient units also outperformed in all sub-indicators except for the *New car emissions*, which present similar levels. However, the gap between the performance of the benchmarks and the inefficient units in *Freight transport* and *GHG emissions* was much more significant. In the *Freight transport* sub-indicator, the inefficient units had a 45% lower share than the one observed in the benchmarks, and in the *GHG emissions* sub-indicator, the inefficient units had on average a value almost 40% higher than the benchmarks. The gap on *Energy dependency*, *Road deaths* and *Public transport* sub-indicators were smaller in this approach, with the inefficient units having an average 10% higher in *Energy dependency*, 3% higher in *Road deaths* and almost 13% lower in *Public transport* than the average of the benchmarks. The results for the *Renewable fuels* share were similar to those obtained with the BoD(K) benchmarks, in which the inefficient units average was only about 60% of the benchmarks' average.

From the above, both models provide similar indications for the inefficient units. It is possible to infer that the majority of the inefficient units' work to improve their sustainability in the transport sector should be focused on reducing the *GHG emissions* from fossil fuel, improving infrastructure and promoting policies to increase the share of *Freight transport* that uses rail and inland waterways and also on increasing the share of *Renewable fuels* in transport. There is also still margin for improvement on the share of *Public transport*, which can benefit from improvements in its accessibility and quality to allow a larger share of passengers to benefit from it, and on *Energy dependency* on oil and petroleum imports, through changes in the transport energy consumption.

To further analyze the difference between the benchmark units of the BoD(K) and the FKHM models, Fig. 8 shows the radar analysis comparing the performance of both groups.

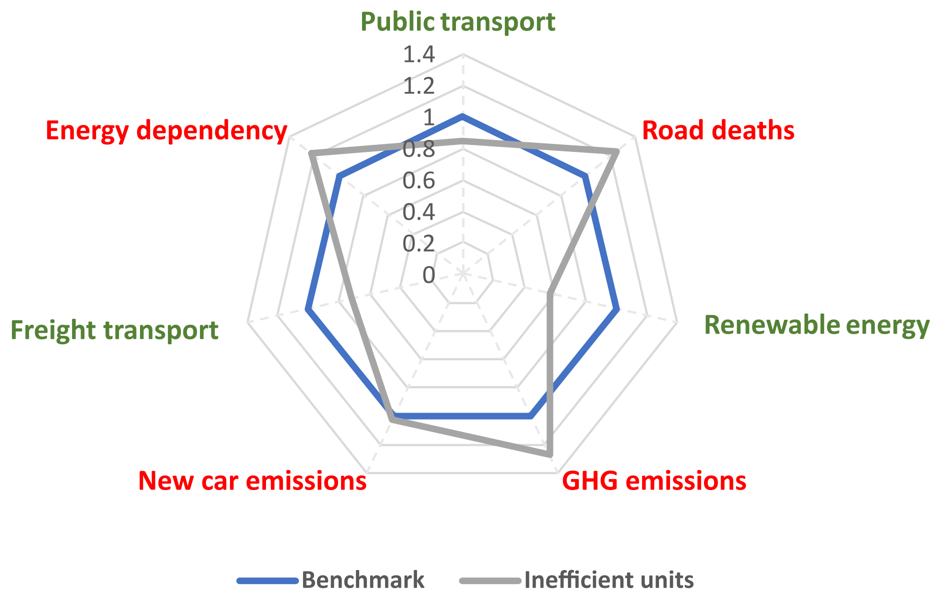

Fig. 6. Sub-indicators comparison between efficient units (benchmarks) and inefficient units from the BoD(K) model.

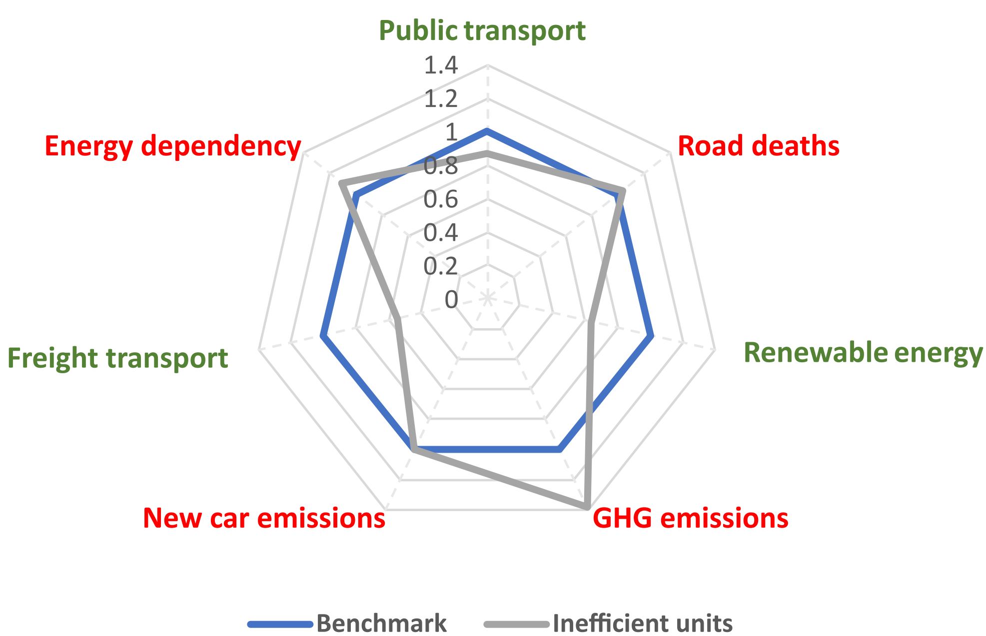

Fig. 7. Sub-indicators comparison between efficient units (benchmarks) and inefficient units from the FKHM model.

The average between both groups on the *Public transport* and *New car emissions* sub-indicators were very similar. Analyzing the forward sub-indicators, the benchmarks from the BoD(K) model have an average 11% higher on *Renewable fuels* than the benchmarks of the FKHM model, but 20% less in the *Freight transport* sub-indicator. In the anti-isotonic sub-indicators, the benchmarks from the BoD(K) model had 16% less *Road deaths*, 8% more *GHG emissions* and 9% less *Energy dependency* than the benchmarks from the FKHM model.

Both profiles of the benchmarks achieved in BoD(K) and FKHM models can be interpreted by the decision-makers and stakeholders as alternative stages to achieve the robust best practices of the benchmarks, which are similar in terms of levels of *Public transport* and *New car emissions*.

Sweden (2018), Latvia (2018) and the Netherlands (2018) are efficient in both models. Table 5 shows the contribution percentage of each benchmark to guide each inefficient country against the best practice frontier. These contributions are determined by the dual of the BoD(K) and FKHM models. According to the model BoD(K), Germany has a $CI = 0.83$ and should emulate the best practices observed in Sweden (2018). Following the FKHM model, Germany has a $CI = 0.55$ and should follow the best practices mainly in Hungary (2015) (47%), but also the ones observed in Latvia (2015) (24%) and Sweden (2018) (29%). Germany is closer to the best practices observed in BoD(K) that promote the increase of *Renewable fuels* for transport and the mitigation of *Road deaths*. For the remaining inefficient countries in 2018, the interpretation is straightforward.

Globally, 8 units achieve a higher CI by using the FKHM model than with the BoD(K) model. Among these, Latvia (2015), Romania (2015), and the Netherlands (2017) become efficient units while the inefficient units include Sweden (2015), Latvia (2016), Romania (2016,

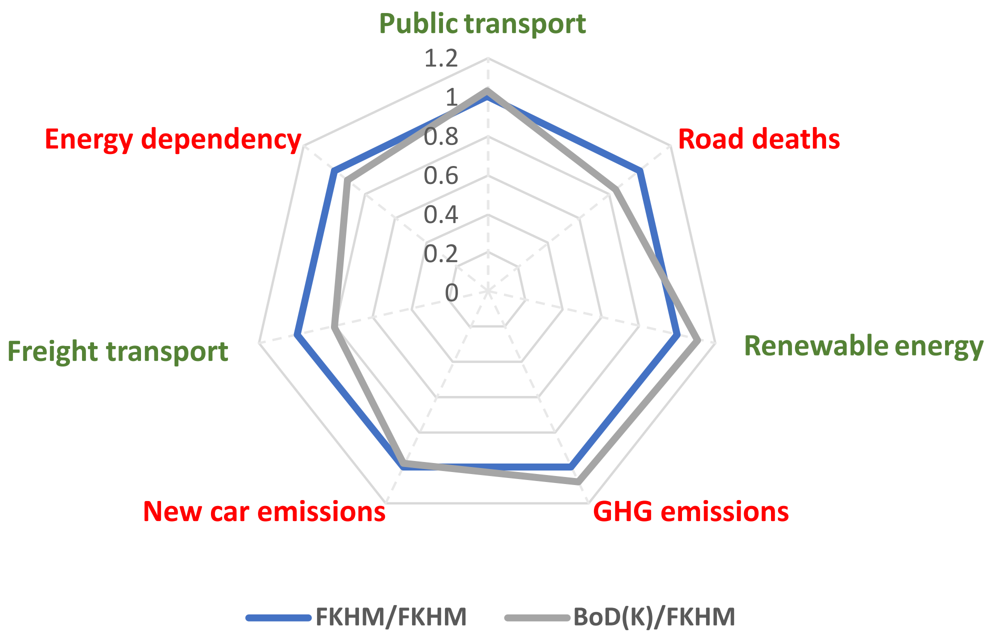

Fig. 8. Sub-indicators comparison between efficient units (benchmarks) from the BoD(K) and FKHM model.

Table 5
Benchmarks of BoD(K) and FKHM models in 2018.

	BoD(K)								FKHM							
	NET15	HUN16	DEN17	HUN17	SWE17	LAT18	NET18	SWE18	LAT15	HUN15	ROM15	DEN17	HUN17	LAT18	NET18	SWE18
BEL	0.00	0.00	0.39	0.12	0.18	0.00	0.00	0.31	0.16	0.69	0.00	0.00	0.00	0.00	0.00	0.15
BUL	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.62	0.11	0.00	0.00	0.00	0.00	0.00	0.28
CZE	0.00	0.79	0.04	0.00	0.00	0.00	0.00	0.17	0.00	0.98	0.00	0.00	0.00	0.00	0.00	0.02
DEN	0.05	0.00	0.76	0.00	0.02	0.00	0.00	0.16	0.00	0.28	0.00	0.67	0.00	0.00	0.00	0.06
GER	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.24	0.47	0.00	0.00	0.00	0.00	0.00	0.29
EST	0.00	0.00	0.00	0.00	0.18	0.00	0.00	0.82	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.79
IRE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GRE	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
ESP	0.00	0.00	0.00	0.00	0.19	0.00	0.61	0.20	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.25
FRA	0.00	0.00	0.33	0.00	0.22	0.00	0.38	0.07	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.33
CRO	0.19	0.00	0.13	0.00	0.00	0.00	0.00	0.69	0.29	0.67	0.00	0.00	0.00	0.00	0.00	0.04
ITA	0.00	0.00	0.00	0.16	0.23	0.00	0.48	0.13	0.00	0.78	0.00	0.00	0.00	0.00	0.00	0.22
CYP	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LAT	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00
LIT	0.00	0.00	0.00	0.00	0.00	0.73	0.00	0.27	0.26	0.00	0.00	0.00	0.00	0.73	0.00	0.01
LUX	0.00	0.45	0.00	0.00	0.55	0.00	0.00	0.00	0.00	0.82	0.00	0.00	0.00	0.00	0.00	0.18
HUN	0.00	0.91	0.05	0.00	0.00	0.00	0.00	0.04	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.04
MAL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NET	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00
AUS	0.04	0.00	0.02	0.50	0.43	0.00	0.00	0.00	0.04	0.66	0.00	0.00	0.00	0.09	0.00	0.21
POL	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.69	0.09	0.85	0.00	0.00	0.00	0.00	0.00	0.06
POR	0.00	0.00	0.90	0.00	0.02	0.00	0.00	0.08	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.52
ROM	0.00	0.00	0.24	0.00	0.24	0.52	0.00	0.00	0.18	0.12	0.65	0.00	0.00	0.00	0.00	0.06
SLN	0.00	0.00	0.00	0.00	0.08	0.06	0.34	0.52	0.54	0.29	0.00	0.00	0.00	0.00	0.00	0.17
SLK	0.00	0.07	0.00	0.57	0.36	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.06	0.00	0.29
FIN	0.00	0.00	0.36	0.00	0.20	0.00	0.11	0.33	0.15	0.26	0.00	0.00	0.00	0.00	0.00	0.59
SWE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**
UK	0.00	0.00	0.61	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.65	0.00	0.00	0.21

2017 and 2018). According to the BoD(K) model, Romania (2018) has a $CI = 0.9209$ and should primarily emulate the best practices of Latvia (2018) (52%), with additional contributions from Denmark (2017) and Sweden (2017) (24%). Under the FKHM model, Romania (2018) achieves a slightly higher $CI = 0.9215$ and should focus on emulating the best practices mainly from Romania (2015) (65%), but also the ones observed in Latvia (2015) (18%), Hungary (2015) (12%), and Sweden (2018) (6%). In this case, Romania (2018) is closer to the best practices frontier identified by the FKHM model and, consequently, prioritizing the adoption of the best practices observed in the benchmarks from this model could provide a more effective pathway for improving performance.

5.2. Discussion

BoD(K) and FKHM models enable to identify nine benchmarks since they are efficient in both models: Denmark (in 2017), Latvia (in 2018), Hungary (in 2015, 2016 and 2017), Netherlands (in 2015 and 2018) and Sweden (in 2017 and 2018). Although, in the FKHM model, Hungary in 2016 and the Netherlands in 2015 and 2018 are only compared with themselves without being benchmarks of inefficient units. From this point of view, BoD(K) has more flexibility in identifying benchmarks.

Specifically, for the inefficient units in both models, if their performance achieved in BoD(K) is higher than the one achieved in the FHKM, it would probably be easier first to pursue the strategy indicated by the BoD(K) since they are closer to their best practices frontier as explained with Germany. The opposite should occur for the five units that have higher performance in the FHKM model than in the BoD(K) model. For the three units that become efficient in the FHKM model and are inefficient in the BoD(K) model, the benchmarks of the BoD(K) should guide them to improve their policies.

Beyond interpreting the composite indicator derived from both models, this paper aims to propose alternative strategies to enhance the sustainability of the global transport sector, regardless of mode (road, rail, waterways, etc.) or segment (freight and passenger). For inefficient units seeking to prioritize improvements in *Renewable fuels*, *Road deaths* and *Energy dependency*, adopting the policies of benchmarks identified by the BoD(K) model is recommended. Alternatively, for those aiming to enhance *Freight transport* and reduce levels of *GHG emissions*, the benchmarks identified by the FKHM model offer more targeted guidance.

It is worth mentioning that the results revealed differences in the average sustainable performance achieved among Eastern, Western, and Nordic European countries, consistent with the findings of Dolge et al. (2023). Despite some variations in the sub-indicators used in both studies, the Nordic European countries present the best performance (an average of $CI = 0.95$ and $CI = 0.82$ on the BoD and FKHM models, respectively) (Dolge et al., 2023), followed by the Eastern European countries (an average of $CI = 0.77$ and $CI = 0.66$ on the BoD and FKHM models, respectively) and the Western European countries, which recorded the lowest averages of $CI = 0.73$ and $CI = 0.51$ on the BoD and FKHM models, respectively.

6. Conclusions

This study aimed to assess the sustainability performance of the transport sector across 28 European countries over a four-year period, aligned with the policy agenda outlined in strategic documents. To assess the sustainability performance, two DEA models have been explored to calculate CI and identify best practices frontiers and their benchmarks.

The methodological approach utilized in this work promoted a comparative analysis of the BoD model proposed by Färe et al. (2019) and the BoD model (Cherchye et al., 2007), using three transformation methods (inversion, constant and MM) for the anti-isotonic sub-indicators. A small, hypothetical, illustrative example was exploited to evaluate these methods. The results of this example revealed slight differences in the CI outcomes across the FKHM and BoD models with the three transformation methods. The best practice frontier was defined by the same efficient units in the FKHM and the BoD-constant models, increasing the robustness of these models from the decision-makers' point of view. Additionally, it is observed that the FKHM model assesses with high efficiency scores the units that have higher levels of forward indicators and, at the same time, have lower levels of reverse sub-indicators. In the case of inefficient units with lower levels of forward and reverse sub-indicators, the FKHM model tends to assess them with lower scores of efficiency. This occurs in this model since the distance from the inefficient unit to the frontier is measured on the observed slope. Thus, the distance between inefficient units and the frontier tends to be higher in the FKHM model than when using BoD-inversion and BoD-constant approaches. These results reinforce the confidence in the robustness and appropriateness of the BoD model with translated anti-isotonic sub-indicators and the FKHM model methods to evaluate the sustainability performance of the transport sector of the 28 countries of the European Union.

Accordingly, the CI for each country and each year was determined by aggregating three forward sub-indicators and four anti-isotonic sub-indicators, chosen to consider the targets for sustainable mobility set in the 2011 White Paper on transport from the European Commission (2011) and the Sustainable Development Goals adopted by the UN (United Nations, 2022).

Comparing the average of CI derived from both models over time, their patterns are similar, although the CI average derived from the FKHM model is almost 20% lower than the one obtained with the BoD(K) model. Nine units were considered robust benchmarks since they are efficient in both models. In this multiple forward and reverse scenario, it is possible to verify that the FKHM model assesses with high efficiency scores the units having higher levels of forward indicators and, at the same time, lower levels of reverse sub-indicators. Similarly, the results achieved in the case study show that the FKHM model tends to decrease the CI score of the inefficient units that have lower scores of forward and reverse sub-indicators.

Based on the results achieved with both models, it is possible to conclude that, in general, the performance of the transport sector in the EU countries has been improving. This result points out that EU countries are making efforts in the right direction, and are strengthening their ability towards sustainability. Additionally, both models give similar indications for the inefficient units to improve their sustainability in the transport sector. Most of the work to improve transport sustainability should be done by reducing GHG emissions from fossil fuel combustion, increasing the share of freight transport that uses rail and inland waterways and also the share of transport energy from renewable sources.

The profiles identified in the benchmarks achieved in both BoD(K) and FKHM models can be interpreted by the decision-makers as alternative stages to achieve the robust best practices of the benchmarks, which are similar in terms of levels of public transport and new car emissions.

Specifically, for the inefficient units in both models, if the performance achieved in BoD(K) exceeds that in FHKM, it would be more effective to first pursue the strategy outlined by the targets proposed by BoD(K), as they are closer to the best practices frontier. This strategy of the transport sector focuses primarily on improving of *Renewable fuels* and levels of *Road deaths* and *Energy dependency*. Conversely, for the fewer units that exhibit higher performance in the FKHM model than in BoD(K), it would likely be more beneficial to follow the strategy suggested by the targets in FKHM. This strategy emphasizes the improvement of *Freight transport* and the reduction of *GHG emissions*.

Funding

This work was supported by the Foundation for Science and Technology (FCT, Portugal), through the Research Centre in Digitalization and Intelligent Robotics (CeDRI) [Grant numbers UIDB/05757/2020 (DOI: 10.54499/UIDB/05757/2020) and UIDP/05757/2020 (DOI: 10.54499/UIDP/05757/2020)], and the Associate Laboratory for Sustainability and Technology in Mountains Regions (SusTEC) [Grant number LA/P/0007/2021 (DOI: 10.54499/LA/P/0007/2020)].

This project is funded by the European Union under Horizon Europe (project 101078933 - STEP - STEM and Equality, Diversity and Inclusion: an open dialogue for research enhancement in Portugal).

CRediT authorship contribution statement

Sarah B. Gruetzmacher: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Clara B. Vaz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ângela P. Ferreira:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the support of Foundation for Science and Technology (FCT, Portugal), CeDRI, SusTEC and STEP project.

Data availability

Data were obtained from Eurostat database, available at https://ec.europa.eu/eurostat/data/database. The dataset was extracted on July 12, 2021.

References

Alper, D., Sinuany-Stern, Z., Shinar, D., 2015. Evaluating the efficiency of local municipalities in providing traffic safety using the data envelopment analysis. Accid. Anal. Prev. 78, 39–50. http://dx.doi.org/10.1016/j.aap.2015.02.014.

Atkinson, A., Cantillon, B., Marlier, E., Nolan, B., 2002. Social Indicators. The EU and Social Inclusion. Oxford University Press, Oxford, http://dx.doi.org/10.1093/0199253498.001.0001.

Babaei, A., Khedmati, M., Jokar, M.R.A., Tirkolaee, E.B., 2023. A decision support framework to evaluate the sustainability performance of urban road transportation. Environ. Sci. Pollut. Res. http://dx.doi.org/10.1007/s11356-023-31265-2.

Charles, V., Färe, R., Grosskopf, S., 2016. A translation invariant pure DEA model. European J. Oper. Res. 249 (1), 390–392. http://dx.doi.org/10.1016/j.ejor.2015.09.037.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European J. Oper. Res. 2 (6), 429–444. http://dx.doi.org/10.1016/0377-2217(78)90138-8.

Chatziioannou, I., Nikitas, A., Tzouras, P.G., Bakogiannis, E., Alvarez-Icaza, L., Chias-Becerril, L., Karolemeas, C., Tsigdinos, S., Wallgren, P., Rexfelt, O., 2023. Ranking sustainable urban mobility indicators and their matching transport policies to support liveable city futures: A MICMAC approach. Transp. Res. Interdiscip. Perspect. 18, 100788. http://dx.doi.org/10.1016/j.trip.2023.100788.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007. An introduction to 'benefit of the doubt' composite indicators. Soc. Indic. Res. 82 (1), 111–145. http://dx.doi.org/10.1007/s11205-006-9029-7.

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. J. Oper. Res. Soc. 59 (2), 239–251. http://dx.doi.org/10.1057/palgrave.jors.2602445.

Chung, W., 2017. Using DEA model without input and with negative input to develop composite indicators. In: 2017 IEEE International Conference on Industrial Engineering and Engineering Management. IEEM, IEEE, pp. 2010–2013. http://dx.doi.org/10.1109/IEEM.2017.8290244.

Cooper, W.W., Seiford, L.M., Tone, K., 2007. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. vol. 2, Springer, New York, http://dx.doi.org/10.1007/978-0-387-45283-8.

Dolge, K., Barisa, A., Kirsanovs, V., Blumberga, D., 2023. The status quo of the EU transport sector: Cross-country indicator-based comparison and policy evaluation. Appl. Energy 334, 120700. http://dx.doi.org/10.1016/j.apenergy.2023.120700.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S., Shale, E.A., 2001. Pitfalls and protocols in DEA. European J. Oper. Res. 132 (2), 245–259. http://dx.doi.org/10.1016/S0377-2217(00)00149-1.

Eisenkopf, A., Burgdorf, C., 2022. Policy measures and their impact on transport performance, modal split and greenhouse gas emissions in german long-distance passenger transport. Transp. Res. Interdiscip. Perspect. 14, 100615. http://dx.doi.org/10.1016/j.trip.2022.100615.

Elisa Fusco, 2015. Enhancing non-compensatory composite indicators: A directional proposal. European J. Oper. Res. 242 (2), 620–630. http://dx.doi.org/10.1016/j.ejor.2014.10.017.

European Commission, 2011. White paper on transport: Roadmap to a single European transport area: Towards a competitive and resource-efficient transport system. URL: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0144:FIN:EN:PDF.

European Commission, 2018a. Directive 2018/2001 of the European parliament and of the council. Off. J. Eur. Union. URL: http://data.europa.eu/eli/dir/2018/2001/oj.

European Commission, 2018b. SHARES tool manual. Unit E.5: Energy. URL: https://ec.europa.eu/eurostat/documents/38154/4956088/SHARES+tool+manual-2019.pdf/.

European Commission, 2019a. Regulation (EU) 2019/631 of the European parliament and of the council. Off. J. Eur. Union. URL: https://eur-lex.europa.eu/eli/reg/2019/631/oj/eng.

European Commission, 2019b. Transport in the European union, current trends and issues. URL: https://transportec.europa.eu/system/files/2019-03/2019-transport-in-the-eu-current-trends-and-issues.pdf.

European Commission, 2020. Sustainable and smart mobility strategy – putting European transport on track for the future. COM/2020/789 final. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789.

European Commission, 2021a. Fit for 55: Delivering the EU's 2030 climate target on the way to climate neutrality. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/delivering-european-green-deal/fit-55-delivering-proposals_en.

European Commission, 2021b. EU energy in figures: statistical pocketbook 2021. Publications Office of the European Union. URL: https://data.europa.eu/doi/10.2833/511498.

European Environment Agency, 2021. Greenhouse gas emissions from transport in Europe, EEA indicator.

Eurostat, 2021. Eurostat database. URL: https://ec.europa.eu/eurostat/data/database. (Accessed 12 July 2021).

Färe, R., Karagiannis, G., Hasannasab, M., Margaritis, D., 2019. A benefit-of-the-doubt model with reverse indicators. European J. Oper. Res. 278 (2), 394–400. http://dx.doi.org/10.1016/j.ejor.2019.02.009,2019.

Fried, H.O., Schmidt, S.S., Lovell, C.K., 1993. The Measurement of Productive Efficiency: Techniques and Applications. Oxford University Press, Oxford, http://dx.doi.org/10.1093/acprof:oso/9780195183528.001.0001.

Fulzele, V., Shankar, R., 2023. Performance measurement of sustainable freight transportation: A consensus model and FERA approach. Ann. Oper. Res. 324 (1), 501–542. http://dx.doi.org/10.1007/s10479-020-03876-.

Georgiadis, G., Papaioannou, P., Politis, I., 2020. Rail and road public transport: Cooperation or coexistence? Transp. Res. Interdiscip. Perspect. 5, 100122. http://dx.doi.org/10.1016/j.trip.2020.100122.

Gruetzmacher, S.B., 2021. Environmental Performance Assessment of the Transport Sector in the European Union (Master's thesis). Instituto Politécnico de Bragança, URL: http://hdl.handle.net/10198/23750.

Gruetzmacher, S.B., Vaz, C.B., Ferreira, A., 2021a. Sustainability performance assessment of the transport sector in European countries. In: Revista Facultad de Ingeniería Universidad de Antioquia. vol. 104, pp. 42–52. http://dx.doi.org/10.17533/udea.redin.20210742.

Gruetzmacher, S.B., Vaz, C.B., Ferreira, A.P., 2021b. Assessing the deployment of electric mobility: A review. In: Gervasi, O., et al. (Eds.), Computational Science and Its Applications – ICCSA 2021. In: Lecture Notes in Computer Science, vol. 12953, Springer, Cham, pp. 350–365. http://dx.doi.org/10.1007/978-3-030-86976-2_24.

Gudmundsson, H., Regmi, M.B., 2017. Developing the sustainable urban transport index. URL: https://api.semanticscholar.org/CorpusID:173172422.

Hoffmann, A., Giovannini, E., 2008. Handbook on Constructing Composite Indicator: Methodology and Users Guide. OECD, Paris, http://dx.doi.org/10.1787/9789264043466-en.

Joumard, R., Gudmundsson, H., 2010. Indicators of environmental sustainability in transport: an interdisciplinary approach to methods. URL: http://cost356.inrets.fr/.

Kutty, A.A., Kucukvar, M., Abdella, G.M., Bulak, M.E., Onat, N.C., 2022. Sustainability performance of European smart cities: A novel DEA approach with double frontiers. Sustainable Cities Soc. 81, 103777. http://dx.doi.org/10.1016/j.scs.2022.103777.

Liu, X., Yuan, M., 2023. Assessing progress towards achieving the transport dimension of the SDGs in China. Sci. Total Environ. 858, 159752. http://dx.doi.org/10.1016/j.scitotenv.2022.159752.

lo Storto, C., Evangelista, P., 2023. Infrastructure efficiency, logistics quality and environmental impact of land logistics systems in the EU: A DEA-based dynamic mapping. Res. Transp. Bus. Manag. 46, 100814. http://dx.doi.org/10.1016/j.rtbm.2022.100814.

Lovell, C.K., Pastor, J.T., Turner, J.A., 1995. Measuring macroeconomic performance in the OECD: A comparison of European and non-European countries. European J. Oper. Res. 87 (3), 507–518. http://dx.doi.org/10.1016/0377-2217(95)00226-X.

Melyn, W., Moesen, W., 1991. Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available. In: Public Economics Research Paper, KUL, Leuven, URL: https://books.google.pt/books?id=2vi9tAEACAAJ.

Morais, P., Camanho, A.S., 2011. Evaluation of performance of European cities with the aim to promote quality of life improvements. Omega 39 (4), 398–409. http://dx.doi.org/10.1016/j.omega.2010.09.003.

Nikolaou, P., Dimitriou, L., 2018. Evaluation of road safety policies performance across Europe: Results from benchmark analysis for a decade. Transp. Res. A 116, 232–246. http://dx.doi.org/10.1016/j.tra.2018.06.026.

Nkesah, S.K., 2023. Making road freight transport more sustainable: Insights from a systematic literature review. Transp. Res. Interdiscip. Perspect. 22, 100967. http://dx.doi.org/10.1016/j.trip.2023.100967.

Reisi, M., Aye, L., Rajabifard, A., Ngo, T., 2014. Transport sustainability index: Melbourne case study. Ecol. Indic. 43, 288–296. http://dx.doi.org/10.1016/j.ecolind.2014.03.004.

Saaty, T., 1980. The Analytic Hierarchy Process. McGraw Hill, New York.

Scheel, H., 2001. Undesirable outputs in efficiency valuations. European J. Oper. Res. 132 (2), 400–410. http://dx.doi.org/10.1016/S0377-2217(00)00160-0.

Shen, Y., Hermans, E., Bao, Q., Brijs, T., Wets, G., 2013. Road safety development in Europe: A decade of changes (2001–2010). Accid. Anal. Prev. 60, 85–94. http://dx.doi.org/10.1016/j.aap.2013.08.013.

Stefaniec, A., Hosseini, K., Assani, S., Hosseini, S.M., Li, Y., 2021. Social sustainability of regional transportation: An assessment framework with application to EU road transport. Socio-Econ. Plan. Sci. 78, 101088. http://dx.doi.org/10.1016/j.seps.2021.101088.

Tsakalidis, A., Boelman, E., Marmier, A., Gkoumas, K., Pekar, F., 2021. Horizon scanning for transport research and innovation governance: A European perspective. Transp. Res. Interdiscip. Perspect. 11, 100424. http://dx.doi.org/10.1016/j.trip.2021.100424.

United Nations, 2022. Sustainable development goals. https://www.un.org/sustainabledevelopment/sustainable-development-goals/. (Accessed 16 August 2022).

United Nations Department of Economic and Social Affairs, 2011. Sustainable transport evaluation: Developing Practical tools for Evaluation in the Context of the CSD Process. Prepared by Daniel Bongardt, Dominik Schmid, Cornie Huizengaand and Todd Litman. In: Partnership on Sustainable, Low Carbon Transport (SLoCaT). Eschborn, Germany, March 2011. Commission on Sustainable Development, Nineteenth Session, 2–13 May 2011. Background Paper No. 10..

Vidoli, F., Fusco, E., Mazziotta, C., 2015. Non-compensability in composite indicators: A robust directional frontier method. Soc. Indic. Res. 122, 635–652. http://dx.doi.org/10.1007/s11205-014-0710-y.

Wong, Y.-H., Beasley, J., 1990. Restricting weight flexibility in data envelopment analysis. J. Oper. Res. Soc. 41 (9), 829–835. http://dx.doi.org/10.1057/jors.1990.120.

Zanella, A., Camanho, A.S., Dias, T.G., 2015a. The assessment of cities' livability integrating human wellbeing and environmental impact. Ann. Oper. Res. 226, 695–726. http://dx.doi.org/10.1007/s10479-014-1666-7.

Zanella, A., Camanho, A.S., Dias, T.G., 2015b. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. European J. Oper. Res. 245 (2), 517–530. http://dx.doi.org/10.1016/j.ejor.2015.03.036.

Zhang, Q., Gu, B., Zhang, H., Ji, Q., 2023. Emission reduction mode of China's provincial transportation sector: Based on "Energy+" carbon efficiency evaluation. Energy Policy 177, 113556. http://dx.doi.org/10.1016/j.enpol.2023.113556, URL: https://www.sciencedirect.com/science/article/pii/S0301421523001416.

Zhou, P., Ang, B., Poh, K., 2007. A mathematical programming approach to constructing composite indicators. Ecol. Econom. 62 (2), 291–297. http://dx.doi.org/10.1016/j.ecolecon.2006.12.020.