European Journal of Operational Research 318 (2024) 1000–1013

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Performance evaluation using multi-stage production frameworks: Assessing the tradeoffs among the economic, environmental, and social well-being

Yiran Niu [a], Jean-Philippe Boussemart [b], Zhiyang Shen [b,*], Michael Vardanyan [b]

[a] The Hong Kong Polytechnic University, Hong Kong
[b] IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management F-59000, Lille, France

ARTICLE INFO

Keywords:
Data envelopment analysis
Environmental technology
Social benchmarking
Nonparametric Estimation

ABSTRACT

Aiming to achieve sustainable development, a constantly growing number of countries have strived to promote economic growth while simultaneously mitigating environmental degradation and maximizing social welfare. However, despite the importance attributed to social well-being in contemporary discourse, its role has not received much attention in the performance evaluation literature. We propose a novel, multi-stage framework based on three dimensions of performance allowing us to assess the tradeoffs between the economic, environmental, and social efficiency in 28 OECD member countries from 2000 to 2019. We construct several scenarios representing policymakers' preferences by altering the weights assigned to the different performance pillars, allowing us to assess the environmental and social repercussions of economic growth. Our findings suggest that policies promoting relatively balanced growth patterns can offer opportunities for higher performance across all three pillars. At the same time, prioritizing development along any single dimension can trigger a relatively significant drop in progress in terms of the other two pillars. We also demonstrate that the sustainable development potential has varied across time and space. Comparisons suggest that the European OECD member countries have outperformed their non-European counterparts in terms of the economic performance, health outcomes, life expectancy, and carbon dioxide (CO_2) emissions. Our results can provide policymakers with insights into strategies for promoting economic growth that account for sustainable development objectives.

1. Introduction

Studies of performance in production have traditionally focused on one-dimensional measures such as those used to approximate the level of economic activity. Earlier attempts to assess economic performance ignore the other important metrics used to define sustainable development, such as environmental degradation and change in human capital. As demonstrated by Chung et al. (1997) and Färe et al. (2012), failure to account for the environmental repercussions of growth by focusing on purely economic performance criteria is likely to yield biased results and therefore lead to wrong policy decisions.

The influential report by Stiglitz et al. (2009) highlighted the limitations of using indicators such as Gross Domestic Product (GDP) to measure performance, acknowledged the importance of pursuing environmentally sustainable growth trajectories, and stressed the need to adopt performance measures that account for the impact of economic activity on the environment. Since material growth often comes at the expense of environmental quality, relying only on GDP as a development criterion can result in disproportionate priorities being placed on economic expansions by policymakers. This could cause resource depletion, environmental degradation and potentially lead to a scenario where a country's economic output could expand without any meaningful corresponding improvement in the living standards of its citizens (Boussemart et al., 2020). The authors of UN's Inclusive Wealth (2018) containing country-level estimates of economic development along with the growth in the natural and human capital find that more than 30 % of the 140 countries studied saw a decrease in their inclusive wealth since late 1990s despite an almost universal increase in their per-capita GDP during the past twenty years (Inclusive Wealth, 2018).

Nordhaus and Tobin (1972) were among the first to offer perspectives on the environmental impact of sustained economic expansions by proposing a measure of environmentally sustainable economic welfare

* Corresponding author.
E-mail address: z.shen@ieseg.fr (Z. Shen).

https://doi.org/10.1016/j.ejor.2024.05.046
Received 30 January 2023; Accepted 28 May 2024
Available online 11 June 2024

that accounts for air pollution levels. At the same time, Leontief (1970, 1973) extended the conventional one-output production framework by allowing for multiple outputs, including unintended byproducts. The notion of environmentally sustainable growth received relatively little attention in the literature until the publication of the United Nations' first report of its Intergovernmental Panel on Climate Change (IPCC, 1990). The report acknowledged the rise in the concentration of greenhouse gases, offered projections for future temperature increases, and highlighted the importance of the resulting impact on society. With a growing focus on environmental degradation in public discourse and greater importance attributed to the notion of environmental sustainability, researchers began to devote progressively more attention to the harmful byproducts generated in production along with socially desirable outputs. For example, Golany et al. (1994) treated sulfur dioxide (SO_2) emissions as a so-called "bad" output in their analysis of Israeli power plants and Lovell et al. (1995) used emissions of carbon dioxide (CO_2) and nitrogen oxides (NOx) to demonstrate that the country-level economic performance changes when these unintended byproducts are used during the assessment of their productive efficiency. Hamilton (1996) incorporated energy consumption and environmental externalities such as pollution into his model of production technology for an open economy.

Attention to green growth has gradually increased thereafter, leading to a large number of studies many of which relied on Data Envelopment Analysis (DEA) to measure economic performance in the presence of environmental externalities. Some of the relatively recent examples include Rakshit and Mandal (2020), who adopted a multi-process network DEA framework to estimate country-level environmental energy efficiency between 1993 and 2013 and compared the differences in the performance of developed, middle-income and developing economies. Shen et al. (2022b) used a multi-stage production approach comprising several sub-technologies to estimate the green productivity growth in China's manufacturing industry and assessed the trade-off between economic and environmental performance. Chen and Chen (2020) estimated an efficiency measure that combines both the economic and environmental aspects associated with the energy-intensive production processes.

However, evolution in human well-being is another important aspect that can be used to measure development and, as such, should be taken into consideration during the assessment of performance along with changes in a country's productive capacity or its environmental conditions. The so-called Better Life Index created by the Organization for Economic Cooperation and Development (OECD) offers a novel framework for measuring social progress and features eleven dimensions of human well-being, including factors such as job security, housing conditions, safety, life satisfaction, as well as the quality of education, environment, and governance, among others (OECD, 2020). According to Stiglitz et al. (2018), aspects such as income equality, employment rate, sound medical services and good education are important determinants of a society's overall level of development in addition to its GDP. The UN's Inclusive Wealth, 2018 defines a nation's wealth as consisting of its economic output but also human and environmental capital (Inclusive Wealth, 2018). Moreover, existing literature suggests that negative externalities represented by socially undesirable outputs can not only cause environmental degradation but also manifest themselves in other domains. More specifically, unintended outcomes can occur in the health care sector as complications from medical procedures or in business, where they can take the form of interest and tax payments for shareholders (Allen, 1999; Smith, 1990).

Hence, the consensus emerging in the literature seems to suggest that any measure of aggregate well-being should account for potential tradeoffs among environmental, social, and economic considerations of performance. However, despite the importance attributed to its role in contemporary discourse, the social dimension has not yet garnered much attention, with most of the existing literature generally focusing on the environmental and economic factors only. Previous studies of performance that account for all three of these aspects simultaneously include Callens and Tyteca (1999), who emphasized the importance of social well-being and environmental performance and used factors such as employment levels, risk of injury, waste generation, and pollution in addition to conventional variables. However, they did not estimate the comprehensive measures of inefficiency they outline in their study. More recently, Boussemart et al. (2020) proposed a DEA framework for assessing multidimensional performance and relied on variables such as the GDP, CO_2, employment levels, health outcomes, and educational attainment to obtain a more inclusive measure of performance. Fukuyama et al. (2020) used a network DEA approach and a sample of Japanese prefectures to estimate a multi-product inefficiency measure that accounts for multiple goals and tradeoffs policymakers face when attempting to promote an increase in environmental, economic, and social commodities. Furthermore, Shen et al. (2022b) measured the productivity of Chinese provinces from 2000 to 2017 by using a framework that combines several dimensions of well-being. Finally, several recent papers have studied the tradeoff among the economic, environmental, and societal considerations of performance in agriculture. For example, Li et al. (2020) propose a framework for optimal land and water allocation under uncertainty that combines the performance objectives defined with respect to these three dimensions, while Kamali et al. (2017) assess the economic, environmental, and social performance of Brazilian soybean farms.

Our analysis contributes to the existing literature in several ways. Given what is still a relatively small number of studies of comprehensive performance that looks beyond an increase in the economic output as a sole objective, we attempt to improve the existing methodology for measuring changes in productivity evaluated across several dimensions of well-being. In addition, while several approaches for including the negative outcomes of economic activity have been proposed in the literature on green performance, existing research has largely ignored the role of socially desirable environmental outcomes that may be generated jointly with the economic output. The framework we propose allows to account for such positive externalities of economic activity. Finally, we contribute to the existing body of knowledge on multidimensional performance by providing insights on the tradeoffs among the different objectives policymakers face when aiming to promote more sustainable development trajectories.

The remainder of the paper is organized as follows. In the next chapter, we provide a short overview of the approaches that have been used to account for the role of socially undesirable outputs. We subsequently introduce our methodology in Section 3, describe the data used in the empirical illustration in Section 4 and discuss the results in Section 5. The last section summarizes our findings and contribution to the literature, as well as offers possible directions for future research along this line of inquiry.

2. Modeling the impact of undesirable outputs

Existing studies of performance incorporating socially undesirable, or so-called "bad," outputs have relied on several approaches for modeling their role in production. The first approach treats bad outputs as freely disposable inputs similar to the conventional production factors such as capital or labor. It has been described in the surveys by Cropper and Oates (1992) and Song et al. (2012) and used by Considine and Larson (2006), Hailu and Veeman (2001), Mandal and Madheswaran (2010), and Mahlberg and Sahoo (2011), among others. Although this method allows to model a production technology in a relatively straightforward fashion, it has been criticized by Färe and Grosskopf (2004, 2006) and Färe et al. (2007), who argue that treating unintended byproducts as inputs is tantamount to allowing for their unlimited increase during production. Furthermore, Färe and Grosskopf (2003) claim that this approach is inconsistent with physical laws while Pethig (2003, 2006) demonstrate that it fails to satisfy the materials balance principle, which was first incorporated into a production framework by

Ayres & Kneese (1969) and posits that the weight of outputs must equal that of inputs.

The second strand of literature treats desirable and undesirable outputs as joint products. Following Shephard (1953) and Shephard and Färe (1974), this approach is based on the notion of null-jointness, which assumes that bad outputs can be completely eliminated only if no desirable outputs are produced. It also assumes the so-called weak disposability between the two categories of outputs, which implies that any decrease in undesirable byproducts at the frontier of technology must be accompanied by a simultaneous reduction in desirable outputs. Early attempts to use this model include Chung et al. (1997), Färe and Grosskopf (1983) and Färe et al. (1986, 1989), and while its more recent applications are described in Dakpo et al. (2016), Färe et al. (2017), and Pham and Zelenyuk (2019). However, the single-equation approach based on the assumption of weak-disposability contradicts the first law of thermodynamics because it neglects the materials-balance condition (Hoang & Coelli, 2011). Similar to the models that treat bad outputs as inputs, this method was also criticized in the literature, first by Førsund (1998, 1999) and later by Russell and Murty (2002) and Murty et al. (2012). As argued by Førsund (2009), it is incapable of accommodating relatively complex relationships that may exist among the variables describing a production process due to its reliance on a unique production function to model different components and stages of this process. An example of this inflexibility is energy generation, which is considered desirable within the purely economic dimension of performance due to its role of a driver of economic activity while simultaneously causing pollution within the environmental dimension.

The third group of studies of production with undesirable outputs relies on the materials balance principle to account for their presence in the production process (Coelli et al., 2007; Hampf & Rødseth, 2019). This approach assumes that the two types of outputs are generated simultaneously and uses material flow coefficients to identify input combinations that produce the minimum material inflow required to produce good outputs, rather than treating undesirable outputs as inputs or outputs.

All three of these approaches approximate a single best-practice production frontier and consider economic dimension as the only measure of performance. More recently, a new methodology based on a so-called by-production technology was introduced by Murty et al. (2012). The by-production framework, which was generalized by Murty (2015) and Murty and Russell (2018), divides the overall production process into components representing individual but interrelated sub-processes, each corresponding to a separate best-practice frontier. Under this approach, bad outputs are treated as unintended consequences of a production activity rather than mere byproducts generated jointly with desirable outputs. Recent empirical studies based on this approach include Ray et al. (2018) and Shen et al. (2022a, 2024), while Dakpo et al. (2016) present a critical review of the recent developments in the literature on the modeling of production technologies characterized by the production of socially undesirable outputs.

3. Methodology

In the production economics literature, the concept of efficiency traces its origins to the pioneering work of Debreu (1951) and Farrell (1957). It is a normative measure that used to assess the performance of decision-making unit (DMU) and can sometimes be interpreted as a ratio of the optimal to actual quantities of inputs and/or outputs. These optimal quantities can be obtained by estimating a so-called best--practice frontier of technology, which represents the boundary of the associated set of production possibilities. The (in)efficiency score represents the gap between the evaluated DMU and its efficient benchmark on the frontier and can therefore be interpreted as the DMU's improvement potential. The efficiency measurement literature can be roughly divided into two families based on the approach used to obtain these inefficiency scores. More specifically, parametric methods require placing assumptions on the functional form of the underlying technology and therefore its best-practice frontier, while nonparametric methods require no such a priori assumptions. The latter category is often based on an approach popularized by Charnes et al. (1978), referred to as the Data Envelopment Analysis (DEA), which has been used in a plethora of performance studies involving situations where DMUs must rely on several inputs to produce a vector of outputs. DEA is based on the mathematical programming algorithms known as the activity analysis model, proposed by Koopmans (1953) and Baumol (1958), and can be used to obtain piecewise linear estimates of the underlying technology's best-practice frontier. Since its inception, DEA has been widely used to study efficiency under various scenarios (Camanho & Dyson, 2005; Chen & Ali, 2004; Cook & Zhu, 2007; Cooper, Ruiz & Sirvent, 2009). We choose to rely on DEA due to its flexibility as we prefer to avoid imposing potentially restrictive assumptions on the exact form of the relationship among the variables we will use to measure performance (Coelli et al. 1998; Lovell & Eeckaut, 1993).

3.1. Production technology and benchmarking models

We use countries as DMUs and formulate a multidimensional framework that considers economic, environmental, and societal dimensions of performance as its integral and interrelated components. Each performance pillar is defined using a nonparametric programming algorithm used to model the activity taking place within that dimension. For example, the purely economic dimension is defined using the standard production axioms attributed to Shephard (1953) that yield conventional production possibility sets. The environmental dimension is modeled using the by-production approach of Murty et al. (2012) that imposes costly disposability of socially undesirable byproducts of economic activity. Finally, the societal considerations are addressed using a benchmarking model introduced by Boussemart et al. (2020).

a) General specification of the three-dimensional performance framework

Let *x, y, z*, and *s* denote the vectors of inputs, economic outputs, environmental outputs and social performance measures, respectively. We distinguish between the desirable (z^+) and undesirable (z^-) environmental outputs as well as between the intended (s^+) and unintended (s^-) social performance outcomes. We assume that the set of all production possibilities can be described using the following convex hull:

$$\begin{aligned}
&T = T_{eco} \cap T^+_{env} \cap T^-_{env} \cap T^+_{soc} \cap T^-_{soc}\\
&= \left\{(x,y,z,s) \in R_+^{M+N+P+Q}:\ x \textit{ can produce}(y,z);\ (y,z)\textit{can generate } s.\right\}:\\
&T_{eco} = \left\{(x,y) \in R_+^{M+N} \middle| f_{eco}(x,y) \le 0\right\},\\
&T^+_{env} = \left\{(x,z^+) \in R_+^{M+P_1} \middle| f^+_{env}(x,z^+) \le 0\right\},\\
&T^-_{env} = \left\{(x,z^-) \in R_+^{M+P_2} \middle| f^-_{env}(z^-,x) \le 0\right\},\\
&T^+_{soc} = \left\{(y,z^+,s^+) \in R_+^{N+P_1+Q_1} \middle| f^+_{soc}(y,z^+,s^+) \le 0\right\},\\
&T^-_{soc} = \left\{(z^-,s^-) \in R_+^{P_2+Q_2} \middle| f^-_{soc}(s^-,z^-) \le 0\right\}.
\end{aligned} \tag{1}$$

where the purely economic sub-technology $T_{eco}$and the two environmental sub-technology sets $T^-_{env}$and T^+_{env} are defined using input quantities x that are assumed to produce desirable economic outputs y, undesirable environmental outputs z^-, and desirable environmental outputs z^+. We measure the purely economic dimension of performance using only the socially desirable outcomes of economic activity and include the undesirable byproducts of this activity in the environmental sub-technology set T^-_{env}. Given their harmful nature, we assume that reducing the level of these bads is costly. Our framework can also account for possible improvements in the environmental conditions by considering additional socially desirable outputs, such as biodiversity indicators, denoted by z^+. These outputs are generated in the sub-technology T^+_{env} and are assumed to be freely disposable. Hence, the continuously differentiable functions that can be used to model these two environmental sub-processes, denoted respectively by f_{env}^{-} and f_{env}^{+}, are characterized by markedly different disposal characteristics. Similarly, the sub-processes modeling the social dimension of performance, denoted by $T_{soc}^{-}$and T_{soc}^{+}, assume that the economic and environmental outputs can have either a negative or a positive impact on the outcomes approximating human well-being. We assume that such outcomes are generated using the production functionsf_{soc}^{-} and f_{soc}^{+} and denote them by s^{-} and s^{+}, respectively. Moreover, *M, N, P*, and *Q* represent the total number of inputs, economic outputs, environmental outcomes, and indicators of social well-being, respectively. The vector *P* contains both desirable (P_1) and undesirable (P_2) outputs, and the vector *Q* consists of both positive (Q_1) and negative (Q_2) social indicators.

Each of these sub-processes is illustrated in Fig. 1. For example, the production function that models the purely economic dimension satisfies the condition $f_{eco}(x,y) \leq 0$ and assumes free, or strong, disposability of inputs. The environmental sub-technology of Murty et al.'s (2012) by-production model imposes costly disposability of the undesirable environmental outputs via $f_{env}^{-}(z^{-},x) \leq 0$, which allows to curb the volumes of undesirable outputs by reducing input use and simultaneously producing lower quantities of desirable outputs. While it is common to account for the negative outcomes of economic activity when evaluating green performance, existing research has largely ignored the role of socially desirable environmental outcomes that may be generated jointly with the economic output. Examples of such positive externalities include higher biodiversity levels, which can be achieved through the spread of innovative agricultural practices. We account for the positive role of these outputs within the environmental sub-technology and, similar to the economic output, assume that they satisfy free disposability via $f_{env}^{+}(x,z^{+}) \leq 0$. Finally, both the economic and environmental outcomes generated within the first two sub-technologies are assumed to have either a positive or negative impact on the social performance indicators. For instance, while harmful byproducts like pollution clearly reduce human well-being, the desirable outputs can potentially have an opposite effect on social welfare. Consequently, the social dimension is modeled using two different processes, given by $f_{soc}^{-}(s^{-},z^{-}) \leq 0$ and $f_{soc}^{+}(y,z^{+},s^{+}) \leq 0$, respectively.

b) Reduced form of the three-dimensional performance framework

The general methodological framework described in the previous section is based on three sub-processes associated with different dimensions of performance. Its general representation yields five separate best-practice frontiers given the additional benchmark models nested within two of these dimensions. However, in our empirical illustration we adopt a more conventional approach for defining the economic and environmental performance pillars. More specifically, the sub-technology T_{eco} models the process whereby GDP is produced using inputs such as labor, capital, and energy. The sub-technologyT_{env}^{-} assumes that energy consumption leads to negative externalities in the form of two harmful byproducts, namely CO_2 emissions and air pollution from fine particulate matter ($PM_{2.5}$). Since total energy in our example is partially derived from fossil fuels and thus cannot be a source of positive environmental externalities, the sub-technology T_{env}^{+} is not operationalized in our illustration. Therefore, T_{env}^{-} will be denoted by T_{env} thereafter. As regards our third performance dimension, the sub-process T_{soc}^{+} treats GDP is an input that drives educational and health outcomes while T_{soc}^{-} assumes that detrimental effects of air pollution can inhibit well-being by increasing premature death rates. Fig. 2 illustrates the reduced form of our multi-dimensional performance framework.

The surface of the production possibilities set *T* can be interpreted as the overall best-practice performance frontier, while the production (in) efficiency can be defined using the distance from individual DMUs inside this set to their associated frontier counterparts. Hence, we can formulate the purely economic production technology T_{eco}, the environmental technology T_{env}, and the benchmarking models T_{soc}^{+} and T_{soc}^{-} that approximate social well-being as

$$\begin{aligned}
T_{eco} &= \left\{(K,L,EGY,GDP) \in R_{+}^{4} \middle| f_{eco}(K,L,EGY;GDP) \leq 0\right\}, \\
T_{env} &= \left\{(EGY,CO_2,PM_{2.5}) \in R_{+}^{3} \middle| f_{env}(CO_2,PM_{2.5};EGY) \leq 0\right\}, \\
T_{soc}^{+} &= \left\{(GDP,EDU,HEA,LIF) \in R_{+}^{4} \middle| f_{soc}^{+}(GDP;EDU,HEA,LIF) \leq 0\right\}, \\
T_{soc}^{-} &= \left\{(PM_{2.5},CO_2,MOR) \in R_{+}^{3} \middle| f_{soc}^{-}(MOR;PM_{2.5},CO_2) \leq 0\right\}.
\end{aligned} \tag{2}$$

where the inputs *K* and *L* that denote capital stock and labor supply, respectively, are assumed to produce *GDP* within the economic sub-technology. The third driver of economic output, or energy use (*EGY*), is a major cause of air pollution and therefore plays an important role within both the economic and environmental sub-technology. While energy consumption impacts the output within T_{eco}, it is also assumed to be the source of negative externalities such as the emissions of CO_2 and anthropogenic sources of $PM_{2.5}$ within T_{env}. Therefore, the amount of energy consumed by these two sub-processes must be equivalent. Our model assumes that when countries use more energy to increase *GDP*

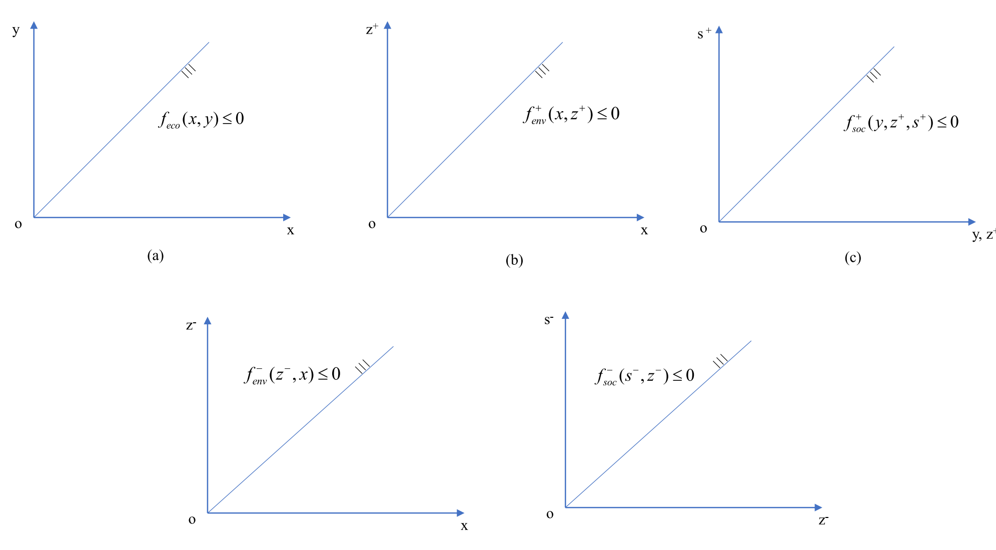

Fig. 1. Sub-processes included in the three-dimensional model of performance.

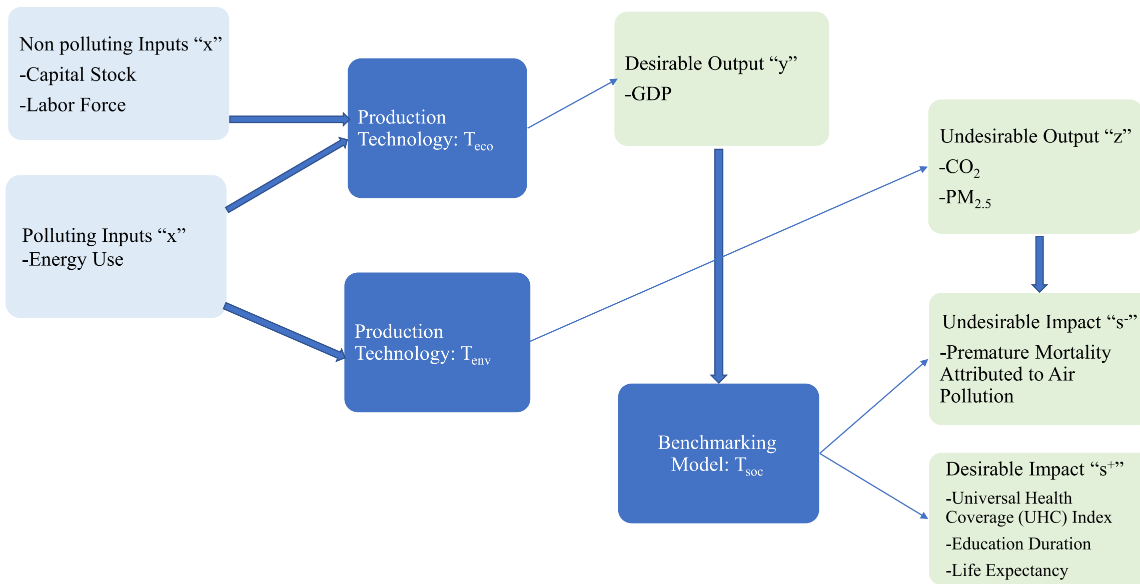

Fig. 2. Reduced form of the multi-dimensional performance framework.

their pollution levels will rise and inhibit environmental performance, underlining the tradeoff between the economic and environmental measures of well-being. Finally, the last two benchmark models allow us to measure performance across the social dimension. The first one relies on the *GDP* as a driver of social well-being indicators such as the level of educational attainment (*EDU*), extent of health coverage (*HEA*), and average life expectancy (*LIF*). The second benchmark measuring social performance assumes that higher levels of CO_2 and $PM_{2.5}$ can cause an increase in premature mortality (*MOR*) without additional investment in pollution abatement technology.

In addition to the variables outlined above, other factors describing economic, environmental, and social well-being can be added to this framework depending on the policymakers' specific objectives and data availability. For example, mortality caused by poor water quality, particularly due to insufficient economic resources and inadequate quality of water supply networks characteristic of some countries, could be included as an additional undesirable outcome within the social performance pillar. Our choice of the above variables is motivated by the fact that our empirical illustration is based on the data from OECD member countries with relatively developed economies. Also, we chose to focus on these measures as they are traditionally used in the literature on multidimensional performance and are readily available from the World Bank through its World Development Indicators database. For example, a number of existing studies of environmental performance treat CO_2, SO_2, and NOx emissions as harmful byproducts of economic activity (Golany et al., 1994; Hampf, 2018; Levkoff, 2011; Lovell et al., 1995; Yang & Pollitt, 2010). However, attempts to account for inhalable particulate matter have so far been limited despite significant health risks attributed to high concentrations of suspended particles due to their capacity to penetrate the respiratory tract.

Given the nature of disamenities such as CO_2 and particle pollution, efforts to reduce their levels are likely to necessitate a reduction in intended outputs such as GDP. By contrast, SO_2 emissions can be curbed by creating incentives for coal-fired power plants to invest in the desulfurization technology, which allows to offset over 90 percent of this gas without incurring any substantial decreases in power generation. Consequently, the assumptions of weak disposability and null-jointness can be violated in situations where byproducts such as SO_2 are generated jointly with the desirable outputs (Dakpo et al., 2016). As regards the social performance measures, the OECD Better Life Index can be used to provide valuable guidance for selecting the indicators of human well-being (OECD, 2020). For example, premature mortality attributed to air pollution serves as a practical measure of a decrease in social welfare that is indirectly linked to high concentrations of air pollutants. While CO_2 and $PM_{2.5}$ are not direct causes of death, their elevated concentrations can contribute to an increase in mortality by exacerbating respiratory and cardiovascular conditions of vulnerable populations. Consequently, a number of countries categorize them as disamenities that can have a negative impact on social well-being.

We rely on a nonparametric linear programming framework to operationalize our model. Following Murty et al. (2012) and Førsund (2018), the economic and environmental technology can be defined as

$$T_{eco} = \left\{ \begin{array}{l} (K, L, EGY, GDP) \in R_+^4 \mid \sum_{i=1}^{I} \lambda_1^i K^i \le K, \sum_{i=1}^{I} \lambda_1^i L^i \le L, \\ \sum_{i=1}^{I} \lambda_1^i EGY^i \le EGY, \sum_{i=1}^{I} \lambda_1^i GDP^i \ge GDP, \ \lambda_1^i \ge 0 \end{array} \right\}, \tag{3}$$

and

$$T_{env} = \left\{ \begin{array}{l} (EGY, CO_2, PM_{2.5}) \in R_+^3 \mid \sum_{i=1}^{I} \lambda_1^i EGY^i = \sum_{i=1}^{I} \lambda_2^i EGY^i, \\ \sum_{i=1}^{I} \lambda_2^i {CO_2}^i \le CO_2, \sum_{i=1}^{I} \lambda_2^i PM_{2.5}^i \le PM_{2.5}, \ \lambda_2^i \ge 0 \end{array} \right\}, \tag{4}$$

where λ_1^i and λ_2^i are vectors of activity variables corresponding to the economic and environmental technology, respectively, i denotes any particular DMU and I is the total number of observations. Besides identifying the best-practice frontiers, they are used to establish the relationship between the two sub-processes via their common component, or energy use, as well as impose assumptions regarding returns to scale. Similarly, performance along the two social axes can be assessed using the models given by

$$T_{soc}^{+} = \left\{ \begin{array}{l} (GDP, EDU, HEA, LIF) \in R_+^4 \mid \sum_{i=1}^{I} \lambda_3^i GDP^i = \sum_{i=1}^{I} \lambda_1^i GDP^i, \\ \sum_{i=1}^{I} \lambda_3^i EDU^i \ge EDU, \sum_{i=1}^{I} \lambda_3^i HEA^i \ge HEA, \sum_{i=1}^{I} \lambda_3^i LIF^i \ge LIF, \ \lambda_3^i \ge 0. \end{array} \right\}, \tag{5a}$$

and

$$T_{soc}^{-} = \left\{ \begin{array}{l} (PM_{2.5}, CO_2, MOR) \in R_+^3 | \sum_{i=1}^{I} \lambda_4^i {CO_2}^i = \sum_{i=1}^{I} \lambda_2^i {CO_2}^i, \\ \sum_{i=1}^{I} \lambda_4^i PM_{2.5}^i = \sum_{i=1}^{I} \lambda_2^i PM_{2.5}^i, \sum_{i=1}^{I} \lambda_4^i MOR^i \leq MOR, \ \lambda_4^i \geq 0 \end{array} \right\}. \tag{5b}$$

where λ_3^i and λ_4^i define the reference technologies associated with the generation of socially desirable outputs and unintended disamenities, respectively. Similar to the assumption that energy used to produce economic output should match its level within the environmental sub-technology, we assume that the output that can be produced using T_{eco} is equivalent to the income available to generate social performance indicators using T_{soc}^{+}. Moreover, since the undesirable outputs CO_2 and $PM_{2.5}$ associated with the environmental sub-technology assume the role of disamenity-generating inputs along the pollution-generating dimension of social performance, the levels of both byproducts within T_{soc}^{-} must match their corresponding counterparts from T_{env}. Thus, our multi-dimensional evaluation framework consists of three performance pillars, divided into four distinct but inter-related processes that define the framework's individual stages.

Finally, to account for the differences in size that may exist among the DMUs, the linear programming models (3)–(5) impose constant returns to scale (CRS) on all technologies. Imposing proportionality conditions on the changes in the inputs and outputs of benchmark observations allows us to perform more reliable comparisons among relatively heterogeneous DMUs.

3.2. The directional distance function and efficiency measurement

The directional distance function (DDF) is an efficiency measure that is well-suited for evaluating performance along several dimensions. Introduced by Chambers et al. (1996), it can be used to assess the distance from each DMU's position inside the corresponding technology set to the best-practice frontier. We will rely on a so-called output-oriented DDF, which attempts to increase all intended outputs while reducing the undesirable outcomes for a given level of inputs across all performance dimensions. The DDF is defined as

$$\begin{aligned} & D\left(x, y, z, s^{+}, s^{-}; 0, g_y; g_z; g_{s^+}; g_{s^-}\right) \\ & = \max\left\{\delta \in \Re^{+} : \left(x, \ y + \delta_y \times g_y, \ z - \delta_z \times g_z, \ s^{+} + \delta_{s^+} \times g_{s^+}, \ s^{-} - \delta_{s^-} \times g_{s^-}\right) \in T\right\}, \end{aligned} \tag{6}$$

where the vector $g = \left(0, g_y, g_z, g_{s^+}, g_{s^-}\right)$ specifies the direction for approaching the frontier of the reference technology. Also, δ_y is the inefficiency score that measures the maximum feasible percentage increase in GDP, δ_z measures the highest possible simultaneous percentage decrease in CO_2 and $PM_{2.5}$, δ_{s^+} approximates the improvement in the extent of health coverage, level of educational attainment, and life expectancy, while δ_{s^-} estimates the potential decrease in pollution-induced premature mortality.

Although the DDF given in (6) is nonnegative, its individual components δ_y, δ_z, δ_{s^+}, and δ_{s^-}, can take on any value due to the performance tradeoffs that exist along different dimensions of well-being. More specifically, negative inefficiency along any given dimension is possible as a consequence of continued worsening in the performance along the other dimensions. Conversely, restricting any given inefficiency component to a positive value implies that performance along no other dimension can be sacrificed to boost efficiency within the pillar in question.

3.3. Weight assignment and nonparametric estimation of the DDF

Given the choice among faster economic growth, better environmental protection, and greater social well-being, aggregating individual inefficiency measures into an overall performance measure requires prior knowledge of policymakers' relative preferences for prioritizing specific development patterns. These preferences are incorporated into the model using a vector of weights $W = \{W_{eco}, W_{env}, W_{soc}\}$, assumed to be exogenous.

A number of different scenarios representing decision-makers' specific policy objectives and priorities can be assumed by changing the assumptions placed on these weights. We consider five such scenarios, summarized in Table 1. Following Shen et al. (2024) and Boussemart et al. (2019), we consider both relatively uniform weighing schemes that target balanced performance and several extreme scenarios, which attribute disproportionate importance to any single dimension of well-being. The two weighting schemes belonging to the former category, denoted by "Balanced1" and "Balanced2," attribute relatively similar importance to all three of the performance pillars while allowing for variations in the specific weights assigned to the individual components of each pillar. For example, the first scenario targeting balanced well-being assigns an equal weight of 1/3 to each performance dimension and then uniformly divides that weight among the dimension-specific outputs. By contrast, the second balanced trajectory attributes an equal weight of 1/7 to each possible outcome, yielding the weights of 1/7, 2/7, and 4/7 for the purely economic, environmental, and social dimension of performance, respectively. Such variations in the scenarios targeting relatively balanced development trajectories allow us to account for different policy preferences and translate to different performance outcomes. The remaining weighing scenarios are described in the last three rows of Table 1. Each of them targets a single performance pillar at a time and assumes a zero weight for the other two performance dimensions. For example, the scenario denoted by "Social" targets social well-being by attributing uniform weights to its defining outcomes while simultaneously disregarding both the economic and environmental pillars of performance.

The DDF defined in the previous section can be estimated using parametric or nonparametric methods. We adopt the latter approach, as it does not require any prior knowledge of the parametric structure of the underlying production technology that is likely to be rather complex due to its multi-dimensional nature. Estimating the values of the DDF requires solving the following linear programming model:

Table 1
Different weighing scenarios (%).

Scenario	Economic	Environmental		Social				Total
	w_{eco} GDP	w_{env} CO_2	w_{env} $PM_{2.5}$	w_{soc} HEA	w_{soc} EDU	w_{soc} LIF	w_{soc} MOR	
Balanced1	33	16.7	16.7	8.3	8.3	8.3	8.3	100
Balanced2	14.3	14.3	14.3	14.3	14.3	14.3	14.3	100
Economic	100	0	0	0	0	0	0	100
Environmental	0	50	50	0	0	0	0	100
Social	0	0	0	25	25	25	25	100

$$
\begin{aligned}
&D\left(x,y,z,s^{+},s^{-};0,g_y;g_z;g_{s+};g_{s-}\right)=\max_{\lambda,\delta}(w_{eco}\delta eco+w_{env}\delta_{env}+w_{soc}\delta_{soc})\\
&=w_{eco}\delta_{GDP}+w_{env}\delta_{CO2}+w_{env}\delta_{PM2.5}+w_{soc}\delta_{EDU}+w_{soc}\delta_{HEA}+w_{soc}\delta_{LIF}+w_{soc}\delta_{MOR}\\
&s.t.\sum_{i=1}^{I}\lambda_1^i K^i\le K',\\
&\sum_{i=1}^{I}\lambda_1^i L^i\le L',\\
&\sum_{i=1}^{I}\lambda_1^i EGY^i\le EGY'\\
&\sum_{i=1}^{I}\lambda_1^i GDP^i\ge GDP'+\delta_{GDP}g_{GDP},\\
&\sum_{i=1}^{I}\lambda_1^i EGY^i=\sum_{i=1}^{I}\lambda_2^i EGY',\\
&\sum_{i=1}^{I}\lambda_2^i CO_2{}^i\le CO_2'-\delta_{CO2}g_{CO2},\\
&\sum_{i=1}^{I}\lambda_2^i PM_{2.5}^i\le PM_{2.5}'-\delta_{PM2.5}g_{PM2.5},\\
&\sum_{i=1}^{I}\lambda_3^i GDP^i=\sum_{i=1}^{I}\lambda_1^i GDP',\\
&\sum_{i=1}^{I}\lambda_3^i EDU^i\ge EDU'+\delta_{EDU}g_{EDU},\\
&\sum_{i=1}^{I}\lambda_3^i HEA^i\ge HEA'+\delta_{HEA}g_{HEA},\\
&\sum_{i=1}^{I}\lambda_3^i LIF^i\ge LIF'+\delta_{LIF}g_{LIF},\\
&\sum_{i=1}^{I}\lambda_4^i CO_2{}^i=\sum_{i=1}^{I}\lambda_2^i CO_2{}^i,\\
&\sum_{i=1}^{I}\lambda_4^i PM_{2.5}^i=\sum_{i=1}^{I}\lambda_2^i PM_{2.5}^i,\\
&\sum_{i=1}^{I}\lambda_4^i MOR^i\le MOR'-\delta_{MOR}g_{MOR},\\
&\lambda_1^i\ge 0,\ \lambda_2^i\ge 0,\ \lambda_3^i\ge 0,\ \lambda_4^i\ge 0.
\end{aligned}
\tag{7}
$$

The variables included in the above model are expressed in levels and thus depend on the absolute size of the countries they represent. It is important to account for differences in the labor supply and total population in samples consisting of heterogeneous countries because their size can have an impact on which economic, environmental, and social performance goals are deemed feasible. For example, the social effort required to attain a given level of health or educational outcomes is expected to be greater in countries with relatively large populations. Similarly, CO_2 emissions and fine particle pollution generally depend on country size. Hence, adding size-related controls to the above model allows us to provide additional context to our approach by taking into account some of the determinants of the variables used in estimation. This can be achieved by changing the inequality constraints associated with labor force to equalities within the economic sub-technology and restricting each country's population to be equivalent across the remaining sub-technologies. Following these modifications, model (7) can be rewritten as

$$
\begin{aligned}
&D\left(x,y,z,s^{+},s^{-};0,g_y;g_z;g_{s+};g_{s-}\right)=\max_{\lambda,\delta}(w_{eco}\delta eco+w_{env}\delta_{env}+w_{soc}\delta_{soc})\\
&=w_{eco}\delta_{GDP}+w_{env}\delta_{CO2}+w_{env}\delta_{PM2.5}+w_{soc}\delta_{EDU}+w_{soc}\delta_{HEA}+w_{soc}\delta_{LIF}+w_{soc}\delta_{MOR}\\
&s.t.\sum_{i=1}^{I}\lambda_1^i K^i\le K',\\
&\sum_{i=1}^{I}\lambda_1^i L^i=L',\\
&\sum_{i=1}^{I}\lambda_1^i EGY^i\le EGY'\\
&\sum_{i=1}^{I}\lambda_1^i GDP^i\ge GDP'+\delta_{GDP}g_{GDP},\\
&\sum_{i=1}^{I}\lambda_1^i EGY^i=\sum_{i=1}^{I}\lambda_2^i EGY^i,\\
&\sum_{i=1}^{I}\lambda_2^i CO_2{}^i\le CO_2'-\delta_{CO2}g_{CO2},\\
&\sum_{i=1}^{I}\lambda_2^i PM_{2.5}^i\le PM_{2.5}'-\delta_{PM2.5}g_{PM2.5},\\
&\sum_{i=1}^{I}\lambda_3^i GDP^i=\sum_{i=1}^{I}\lambda_1^i GDP',\\
&\sum_{i=1}^{I}\lambda_3^i EDU^i\ge EDU'+\delta_{EDU}g_{EDU},\\
&\sum_{i=1}^{I}\lambda_3^i HEA^i\ge HEA'+\delta_{HEA}g_{HEA},\\
&\sum_{i=1}^{I}\lambda_3^i LIF^i\ge LIF'+\delta_{LIF}g_{LIF},\\
&\sum_{i=1}^{I}\lambda_4^i CO_2{}^i=\sum_{i=1}^{I}\lambda_2^i CO_2{}^i,\\
&\sum_{i=1}^{I}\lambda_4^i PM_{2.5}^i=\sum_{i=1}^{I}\lambda_2^i PM_{2.5}^i,\\
&\sum_{i=1}^{I}\lambda_4^i MOR^i\le MOR'-\delta_{MOR}g_{MOR},\\
&\sum_{i=1}^{I}\lambda_2^i POP^i=\sum_{i=1}^{I}\lambda_3^i POP^i=\sum_{i=1}^{I}\lambda^i{}_4 POP^i=POP'\\
&\lambda_1^i\ge 0,\ \lambda_2^i\ge 0,\ \lambda_3^i\ge 0,\ \lambda_4^i\ge 0.
\end{aligned}
\tag{8}
$$

where POP^i denotes country i's total population. Compared to the linear program (7), model (8) accounts for the size of countries' population and workforce during the measurement of their performance.

The above specification defines performance using the absolute quantities of outcomes across different dimensions of well-being. Therefore, it is incapable of distinguishing between different measurement scales characterizing its variables and properly accounting for these differences. At the same time, situations involving groups of variables with different measurement scales are likely when a relatively large number of outcomes must be used to measure efficiency. To render comparisons between different categories of outcomes possible, all variables included in the model must be defined as scale-invariant. This can be achieved by defining their absolute values on a per capita basis and relying on ratios to measure efficiency, bringing our approach more in line with the methodology nonprofit organizations such as The World Bank use to measure performance and well-being. As demonstrated below, this adjustment leads to an additional restriction on the activity variables λ^i from model (8) that assumes constant returns to scale (CRS). Ray (2004; 2015) and Ray et al. (2021) demonstrate that a maximization problem containing variables expressed in levels under a CRS technology is strictly equivalent to a specification based on the ratios of these variables that assumes variable returns to scale (VRS).

Expressing the variables defined in levels as quantities per unit of labor in the economic sub-technology yields the following set of restrictions:

Table 2
Variables used in estimation.

Pillar	Technology	Variable	Input/ Output Status	Details and Measurement Units
Economic	Production technology	Capital stock (*K*)	Input	Total capital stock based on PPP (billions of 2017 USD)
		Labor force (*L*)	Input	Number of employees (millions)
		Energy use (*EGY*)	Input	Total energy consumption (quadrillion Btu)
		GDP	Output	Total output based on PPP (millions of 2011 USD)
Environmental	Polluting technology	Population (*POP*)	Input	Total population (tens of millions)
		CO_2 emissions (CO_2)	Output	Millions of tons
		Particle pollution ($PM_{2.5}$)	Output	Population-weighted exposure to ambient $PM_{2.5}$ pollution (micrograms per m^3)
Social	Benchmark models	UHC service coverage index (*HEA*)	Output	Coverage index for essential health services (scale of 0 to 100)
		Education (*EDU*)	Output	Number of years of primary and secondary education (years)
		Life expectancy (*LIF*)	Output	Life expectancy at birth (years)
		Premature mortality (*MOR*)	Output	Number of deaths per 100,000 inhabitants from outdoor and indoor air pollution.

$$\begin{aligned}
&\sum_{i=1}^{I}\lambda_1^i\frac{K^i}{L^i}\le\frac{K'}{L'},\\
&\sum_{i=1}^{I}\lambda_1^i\frac{EGY^i}{L^i}\le\frac{EGY'}{L'},\\
&\sum_{i=1}^{I}\lambda_1^i\frac{GDP^i}{L^i}\le\frac{GDP'}{L'}+\delta_{GDP}g_{GDP}\ \text{with}\ g_{GDP}=\frac{GDP'}{L'},\\
&\sum_{i=1}^{I}\lambda_1^i\frac{L^i}{L^i}=\frac{L'}{L'}\Leftrightarrow\sum_{i=1}^{I}\lambda_1^i=1.
\end{aligned}\tag{9}$$

Similarly, defining the variables modeling the environmental sub-technology and social benchmarks on a per-capita basis allows us to rewrite the associated restrictions in the following fashion:

$$\begin{aligned}
&\sum_{i=1}^{I}\lambda_1^i\frac{EGY^i}{POP^i}=\sum_{i=1}^{I}\lambda_2^i\frac{EGY^i}{POP^i},\\
&\sum_{i=1}^{I}\lambda_2^i\frac{CO_2^i}{POP^i}\le\frac{CO_2'}{POP'}-\delta_{CO2}g_{CO2},\ \text{with}\ g_{CO2}=\frac{CO_2'}{POP'}\\
&\sum_{i=1}^{I}\lambda_2^i\frac{PM_{2.5}^i}{POP^i}\le\frac{PM_{2.5}'}{POP'}-\delta_{CO2}g_{CO2},\ \text{with}\ g_{CO2}=\frac{PM_{2.5}'}{POP'}\\
&\sum_{i=1}^{I}\lambda_3^i\frac{GDP^i}{POP^i}=\sum_{i=1}^{I}\lambda_1^i\frac{GDP'}{POP'},\\
&\sum_{i=1}^{I}\lambda_3^i\frac{EDU^i}{POP^i}\ge\frac{EDU'}{POP'}+\delta_{EDU}g_{EDU},\ \text{with}\ g_{EDU}=\frac{EDU'}{POP'}\\
&\sum_{i=1}^{I}\lambda_3^i\frac{HEA^i}{POP^i}\ge\frac{HEA'}{POP'}+\delta_{HEA}g_{HEA},\ \text{with}\ g_{HEA}=\frac{HEA'}{POP'}\\
&\sum_{i=1}^{I}\lambda_3^i\frac{LIF^i}{POP^i}\ge\frac{LIF'}{POP'}+\delta_{LIF}g_{LIF},\ \text{with}\ g_{LIF}=\frac{LIF'}{POP'}\\
&\sum_{i=1}^{I}\lambda_4^i\frac{CO_2^i}{POP^i}=\sum_{i=1}^{I}\lambda_2^i\frac{CO_2^i}{POP^i},\\
&\sum_{i=1}^{I}\lambda_4^i\frac{PM_{2.5}^i}{POP^i}=\sum_{i=1}^{I}\lambda_2^i\frac{PM_{2.5}^i}{POP^i},\\
&\sum_{i=1}^{I}\lambda_4^i\frac{MOR^i}{POP^i}\le\frac{MOR'}{POP'}-\delta_{MOR}g_{MOR},\ \text{with}\ g_{MOR}=\frac{MOR'}{POP'}\\
&\sum_{i=1}^{I}\lambda_2^i\frac{POP^i}{POP^i}=\sum_{i=1}^{I}\lambda_3^i\frac{POP^i}{POP^i}=\sum_{i=1}^{I}\lambda_4^i\frac{POP^i}{POP^i}=\frac{POP'}{POP'}\Leftrightarrow\sum_{i=1}^{I}\lambda_2^i=\sum_{i=1}^{I}\lambda_3^i=\sum_{i=1}^{I}\lambda_4^i=1.
\end{aligned}\tag{10}$$

Unlike the linear programming models (7) and (8), the specification based on the original objective function and the restrictions in (9) and (10) is scale-invariant because all its variables are defined either per unit of labor input or per capita. This allows us to compare the evolution in the access to health care, expressed as an index whose measurement scale differs from that of the other variables, to the changes in the other outcomes of social well-being or environmental performance such as life expectancy or CO_2 emissions, defined in levels. Expressing our variables on a per capita basis offers a more operational solution given the differences among their associated measurement scales.

4. Data

We operationalize our model using a balanced panel of 28 OECD Member countries for the period 2001–2019.[1] Most of the input and output variables are obtained from The World Bank's World Development Indicators database and the data reported by the US Energy Information Administration.[2] The variables are summarized in Table 2 along with their units of measurement. For example, the value of GDP is measured at purchasing power parity (PPP) and expressed in 2011 US dollars. Total energy consumption is obtained by aggregating energy from fossil fuels, nuclear power, and renewable energy. To account for differences in both work duration and workforce qualification levels across different countries, our labor input is weighted by the average annual employee working time and a human capital index. The latter is calculated using country-level educational attainment levels and the rate of returns to education. As regards the environmental technology, we define particle pollution using exposure to inhalable particulate matter, measured using mean annual concentration of fine particulates and weighted by total population (Cohen et al., 2017). Performance along the social dimension is assessed using metrics such as the level of educational attainment (EDU), defined as the number of years of primary and secondary education. We also rely on the universal health coverage (UHC) service coverage index, retrieved from Our World in Data website, to control for the differences in the access to essential health services across different countries.[3] This performance measure is denoted by HEA in our models. Moreover, life expectancy (LIF) is included among the outcomes of social well-being along with premature mortality (MOR) attributed to both outdoor and indoor air pollution. The latter is calculated as the number of deaths per 100,000 inhabitants and is standardized by assuming a constant age structure of the population to allow for comparisons among different countries and over time.[4]

Recall that our main objective is to assess temporal changes in the environmentally and socially sustainable performance. This can be achieved by using the variables described above and the DDF to estimate country-level efficiency across different dimensions of well-being separately for each year.

The descriptive statistics displayed in Table 3 suggest a relatively high degree of heterogeneity among the countries included in our sample. For example, the values of GDP range from approximately 12,147 million USD in Iceland in 2001 to about 20,507 billion USD in the United States in 2019. Similarly, the highest $PM_{2.5}$ concentration levels were recorded in Japan in 2011 and corresponds to 97.6 micrograms per cubic meter. By contrast, Greece reported particle pollution level of only 5.3 μg per m^3 in 2019. We observe similar differences in the case of pollution-related premature death rates, carbon dioxide emissions, and economic inputs such as capital and labor.

[1] The countries included in the sample are Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Republic of Korea, Slovakia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

[2] See https://data.worldbank.org/indicator and https://www.eia.gov/international/data/world for more details.

[3] The UHC index is reported on a unitless scale of 0 (worst performance) to 100 (best performance) and is computed as a geometric mean of 14 indicators of basic health service coverage. Additional details and the data are available at https://ourworldindata.org/grapher/universal-health-coverage-index.

[4] Data processed by Our World in Data and available from the Global Burden of Disease Study conducted by the Institute of Health Metrics and Evaluation (IHME, 2019).

5. Results

5.1. Performance shortfalls under different scenarios

As mentioned in Sections 3.2 and 3.3, allowing for negative inefficiency along any given performance dimension or changing the weights on the different pillars of well-being is likely to affect the results. We restrict all inefficiency components to be non-negative and focus on the findings under different weighing scenarios. It is noteworthy that since CO_2, $PM_{2.5}$ and MOR are all undesirable from society's perspective, a positive value of their associated inefficiency measures, denoted respectively by δ_{C02}, $\delta_{PM2.5}$, and δ_{MOR} in the objective function, signals a potential for improved performance that can be realized by decreasing these outcomes. Similarly, negative inefficiency scores corresponding to these undesirable outputs suggest possible improvement in the overall performance that can be achieved at the expense of increased pollution and higher mortality.

In Table 4, we summarize the results from the specification assuming non-negative inefficiency, defined as percentage improvement across each outcome. We measure performance under five different scenarios approximating heterogeneous policy preferences described earlier (see Table 1). Recall that the first two settings, whose corresponding results are displayed in the first two rows, attribute relatively similar importance to each performance pillar and their defining outcomes. The results corresponding to the remaining three scenarios, which target a single pillar at a time, are given in the last three rows. Additionally, the second column of Table 4 reports the overall weighted inefficiency under each setting. It is defined as the average overall gap between the observed performance and its feasible best-practice benchmark and can thus be interpreted as the overall improvement potential.

Results from the two specifications assuming balanced weights are relatively similar to each other and suggest that the global performance taking into account the economic, environmental and social well-being can be improved by an average of around 26 % across the OECD countries. Although this improvement is feasible by targeting any of the outcomes included in our model, decreasing undesirable outcomes such as premature mortality, fine particle pollution, and CO_2 emissions looks particularly promising. More specifically, our first weighing scheme suggests an average shortfall of 16 % in GDP, surpluses of respectively 34 % and 43 % in disamenities such as CO_2 and $PM_{2.5}$, and a deficit of 14 % in health service coverage between 2000 and 2019. It also points to an average gap of around 6 % in the education level, a shortfall of about 5 % in life expectancy, and excess mortality of nearly 64 % during the period considered. In general, inefficiency associated with both the economic and social pillar is lower than for the environmental dimension of performance. Apart from significant shortfalls in premature mortality, OECD countries appear to have performed relatively well in terms of their social performance.

These findings add perspective to the results reported in the UN's Inclusive Wealth, 2018, which measures the evolution of countries' multidimensional performance by estimating monetary values that can

Table 3
Descriptive statistics.

Variable	Mean	Std. Dev.	Min	Max
K	7,498,958	12,032,474	57,470.1	69,059,088
L	115,769.5	186,440.4	745.3	1,047,767
EGY	8.4	18.1	0.1	102.5
GDP	1,782,672	3,235,277	12,147.3	20,507,080
POP	43	61.4	0.3	329.1
CO_2	460.4	1024.3	2.5	6015.5
$PM_{2.5}$	20.2	17.8	5.3	97.6
HEA	74.4	11.2	31	89
EDU	12.3	0.6	10	14
LIF	77.6	5.6	53.4	84.4
MOR	39.8	45.8	2.7	242.6

Table 4
Average performance shortfalls assuming non-negative inefficiency under various weighing scenarios (%).

Scenarios	Overall Inefficiency	δ_{GDP}	δ_{C02}	$\delta_{PM2.5}$	δ_{HEA}	δ_{EDU}	δ_{LIF}	δ_{MOR}
Balanced1	25.7	16	34.4	43	14	6.4	5.2	63.6
Balanced2	26.2	14.4	35.9	42.6	14.1	7.2	5.3	64.3
Economic	17.1	17.1	0.1	0.1	0.1	0.1	0.2	0
Environmental	40.4	0.1	37.7	43.2	0.1	0.1	0.1	0.2
Social	23.2	0.0	0.2	0.1	13.8	8.5	5.5	64.9

Table 5
Regional comparisons between performance shortfalls under the first scenario (%).

Overall Inefficiency	δ_{GDP}	δ_{C02}	$\delta_{PM2.5}$	δ_{HEA}	δ_{EDU}	δ_{LIF}	δ_{MOR}
European OECD member countries							
25.4	14.2	31.7	46.8	10.2	6.7	3.7	70.5
Non-European OECD member countries							
26.3	20.6	41.1	33.4	23.7	5.4	8.8	46.3

be assigned to the change in their economic output, human capital, and environmental commodities over time (Inclusive Wealth, 2018). One of the main conclusions of the Inclusive Wealth (2018) is that while countries' economic output and social well-being have both steadily increased over the last twenty years, the value of their environmental assets has declined. When measured in monetary terms, improved social well-being measured as an increase in human capital along with greater economic wealth have exceeded the simultaneous decrease in the natural capital. Our results are in line with these conclusions when performance across the economic, environmental, and social dimension is assessed using relatively similar weights. They highlight environmental performance gaps among the countries as a possible culprit behind the decrease in the natural capital, suggesting that policymakers can help slow down this decrease by promoting better performance within the environmental sub-technology.

In addition, our results offer convincing evidence of significant tradeoffs among the different dimensions of well-being when any single pillar is assumed to take precedence over the other two measures. For example, the extreme case involving economic expansion as the only policy objective suggests an average GDP gap of around 17 % among the countries included in the study. By contrast, both the environmental and social performance can be improved by only 0.1 % under this scenario, suggesting that targeting better economic outcomes leaves little room for progress across the remaining performance dimensions. The weight distribution assumed under this scenario yields the lowest average overall inefficiency among the five weighing scenarios we considered, highlighting the importance attributed to economic growth among the OECD member countries. Similarly, focusing solely on environmental efficiency can help reduce CO_2 emissions and $PM_{2.5}$ concentration levels by approximately 38 % and 43 %, respectively, albeit at the expense of foregoing economic growth and enhanced social well-being, whose associated inefficiency levels are nearly zero under this scenario. It is noteworthy that the setting exclusively targeting environmental performance corresponds to the highest overall inefficiency of around 40 %, which can be interpreted as a sign of a significant slack in environmental performance among the countries included in our sample. Finally, results in the last row of Table 4 suggest that adopting social well-being as the only policy goal is likely to lead to both slower economic development and insufficient environmental protection. Development along any single dimension comes at the expense of slower progress in terms of the other performance pillars, underscoring the importance of adopting a balanced approach to promoting growth by policymakers.

5.2. Regional comparisons and temporal changes

The OECD member countries included in our study share a number of similarities as technologically advanced and economically developed nations known for their relatively high levels of social well-being. In the next stage of our analysis, we assess the role of geographical location in explaining the differences in multidimensional performance and improvement potential under the balanced weighting scenarios. Specifically, we distinguish between the European and non-European OECD member states, with the former category including the majority of the European countries.[5] Many of the European OECD member states are part of important institutional frameworks such as the European Union (EU) and Eurozone, providing additional rationale behind distinguishing between countries based on their geographical location. Additionally, they exhibit significant cultural similarities, possess comparable political systems, and are characterized by a notable level of economic and technological interdependence.

Table 5 highlights the differences in inefficiency along different dimensions between the European and non-European OECD economies under the first balanced weighting scenario. As before, these estimates represent average inefficiency levels measured as performance shortfalls associated with individual outcomes. Despite similarities in the overall performance, European countries appear to have fared better in terms of the purely economic performance, health outcomes, life expectancy, and CO_2 emissions compared to their non-European counterparts. One possible explanation for this is the EU's recent environmental protection initiatives, which helped reduce emissions among the European OECD members between 1990 and 2014 (Inclusive Wealth, 2018). However, the non-European OECD member states outperformed their European counterparts in the areas of fine particle pollution and especially life expectancy, which is consistent with a relatively high intensity of per capita $PM_{2.5}$ damages in Europe reported in the Inclusive Wealth (2018). Apart from the premature death rate, the OECD member countries located in Europe have performed generally better than did the non-European OECD economies.

In table 6, we list the two best and worst performers among the countries included in the study. Switzerland, which is our best performer overall, has fared better than most other countries in terms of nearly all individual components of the three well-being pillars. Despite its relatively poor economic efficiency, Greece is among the benchmark nations defining the best-practice frontier of the environmental sub-technology, while simultaneously performing relatively well in terms of social well-being. As for the Czech Republic, its relatively high overall inefficiency is driven by poor economic and environmental performance. Finally, Japan underperforms in terms of both environmental and social development, especially when it comes to fine particle pollution and access to essential health services.

In addition to comparing different regions and countries, we analyze the temporal changes in performance gaps along the three well-being dimensions under the balanced weighing scenario. We also decompose the evolution in the environmental and social inefficiency into changes in their corresponding components over time. These trends are illustrated in Figs. 3-5.

[5] The 20 European OECD member countries are Austria, Belgium, Czech Republic, Denmark, Germany, Finland, France, Greece, Hungary, Iceland, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, and United Kingdom. The remaining OECD member states include Australia, Canada, Japan, Mexico, New Zealand, Republic of Korea, Turkey, and United States.

Table 6
Comparisons between the best and worst performers under the first scenario.

	Overall Inefficiency	δ_{GDP}	δ_{C02}	$\delta_{PM2.5}$	δ_{HEA}	δ_{EDU}	δ_{LIF}	δ_{MOR}
Switzerland	0	0	0	0	0	0	0	0.5
Greece	14.6	40.6	0	0	3.4	8.8	1.0	0
Czech Republic	59.3	67.0	58.5	87.7	30.8	14.0	9.5	96.7
Japan	58.7	45.1	53.3	91.3	104.8	10.5	22.1	97.3

Looking first at Fig. 3, we can see that the economic inefficiency has declined between 2001 and 2009 before converging to a comparatively steady annual rate of around 14 % during the second half of the period studied. The adverse economic impact of the Great Recession is not evident among these findings, possibly due to the global nature of this downturn that affected a relatively large number of economies, including the benchmark countries that defined the frontier of the economic sub-technology in late 2000s. At the same time, mean performance along the social dimension of well-being has remained relatively steady throughout the entire period considered, which is likely due to the sample structure marked by a high share of developed countries with relatively strong social safety nets. We can also see that the gap between the economic and social pillar appears to have widened over time, highlighting the relative importance OECD member countries attribute to their economic development. As regards the evolution in the environmental efficiency, we can see that this category is the only pillar that shows a declining performance over time. This finding appears to be consistent with the results reported in the Inclusive Wealth (2018), which suggest a global decrease of 0.7 % in the value of countries' natural capital between 1990 and 2014.

Finally, Figs. 4 and 5 describe the temporal changes in the average inefficiency corresponding to the components of the environmental and social pillar. We note that performance gaps associated with particle pollution have not only remained relatively high but also increased during the last decade, helping explain the overall decrease in the environmental efficiency mentioned earlier. At the same time, differences in country-level performance measured in terms of CO_2 emissions appear to have slightly narrowed over time, despite remaining relatively high at more than 30 % in 2019, on average. This result aligns with a generally slower growth rate in CO_2 emissions among the developed countries than worldwide observed since 1990 (OECD, 2023). As far as the social dimension of well-being is concerned, we note a steady decline in the mean inefficiency measured in terms of access to essential health care. This favorable trajectory includes a sharply improved performance beginning in 2010, which can be attributed to the signing of the U.S. Affordable Care Act in May of that year and subsequent improvement in the access to health care in the U.S. By contrast, performance gaps related to premature death rates from air pollution have remained relatively high during the last two decades and appear to be widening with time.

6. Conclusion and policy implications

Promoting progress along multiple dimensions of performance is an important challenge for any government looking to promote sustainable growth. Ambitious economic performance objectives that are frequently

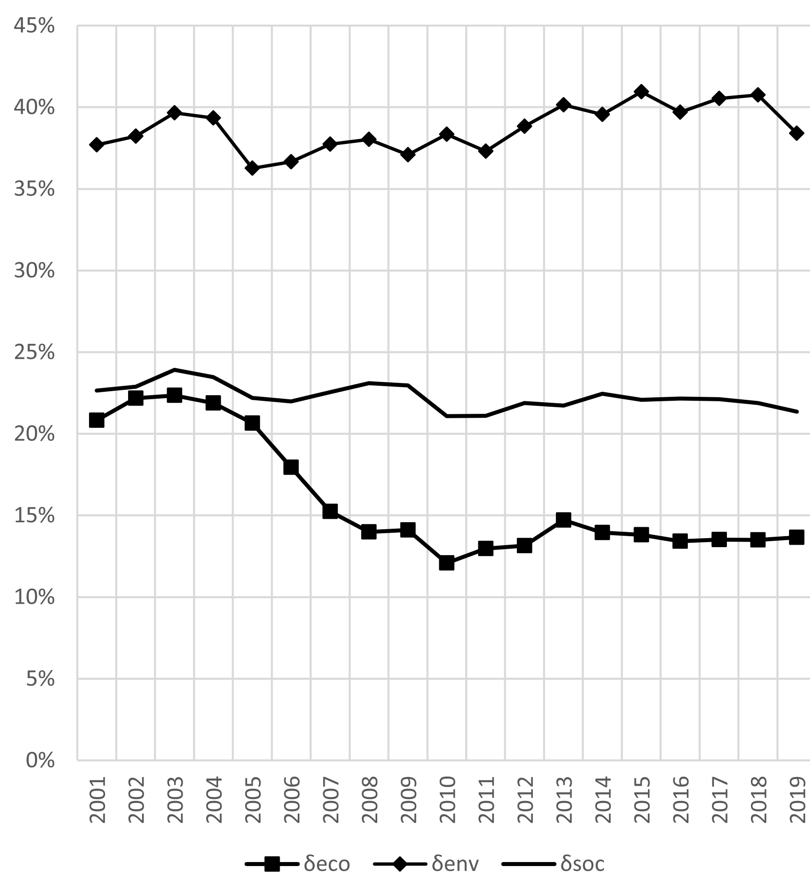

Fig. 3. Temporal changes in different performance pillars under the first scenario.

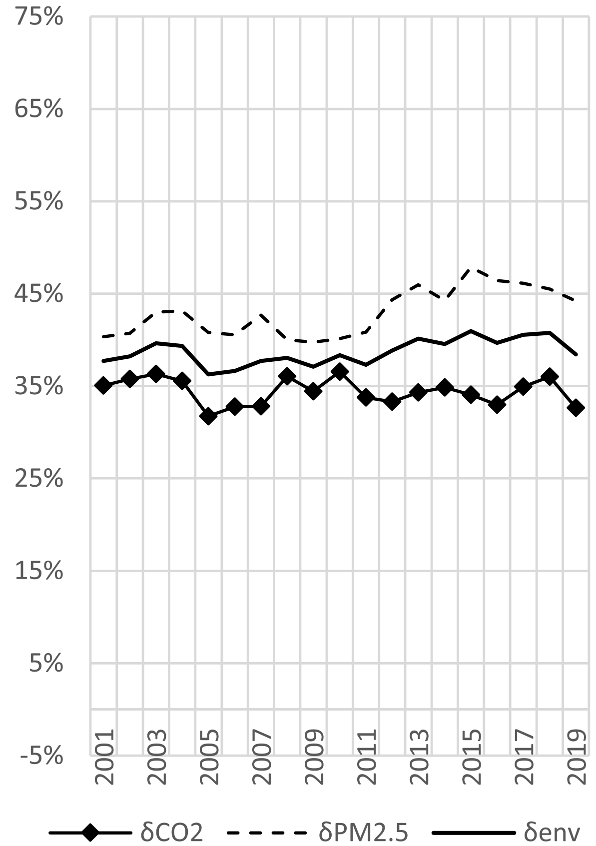

Fig. 4. Decomposition of the environmental performance pillar under the first scenario.

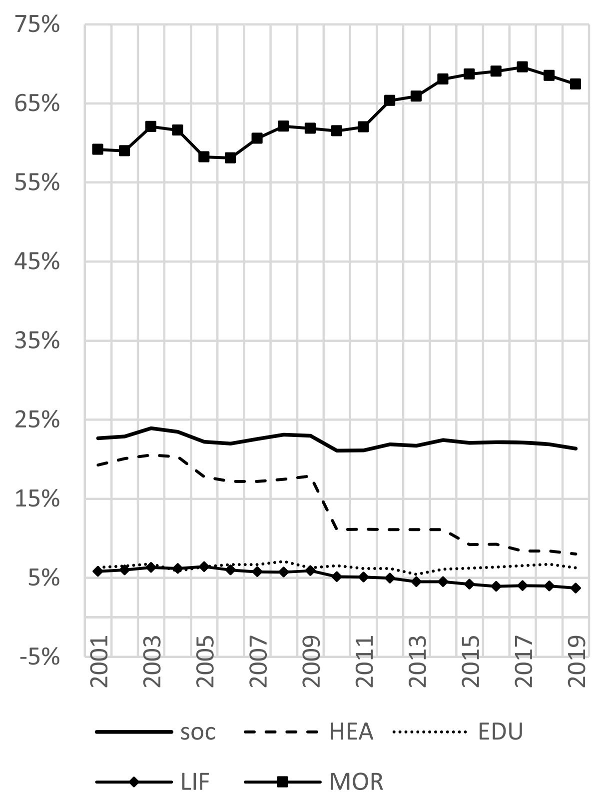

Fig. 5. Decomposition of the social well-being pillar under the first scenario.

prioritized by policymakers carry environmental repercussions and impose social costs, emphasizing the tradeoffs among the growth in economic output, environmental protection, and social progress. Our approach accounts for these tradeoffs, allowing us to measure country-level performance along multiple dimensions of well-being simultaneously. More specifically, we define multidimensional performance measures by focusing on three pillars of sustainability, including economic activity, environmental protection, and social well-being.

We measure performance using the by-production approach of Murty et al. (2012), which treats the production of intended outputs and generation of disamenities as separate but interconnected parts of a model of economic performance in the presence of environmental externalities. Building on the work of Ray et al. (2018) and Boussemart et al. (2020), we extend this framework by adding the social dimension of well-being to the by-production model, defined using outcomes such as access to health care, education level, life expectancy, and death rate from pollution. We rely on both the intended and unintended manifestations of social performance to define the social performance pillar and assume that it is closely related to both the economic and environmental technology. In addition to using CO_2 as a measure of environmental degradation, we add pollution from fine particulates to the sub-process that models environmental performance. The EU's European Environment Agency (2022) estimates that exposure to abnormally high $PM_{2.5}$ concentration levels resulted in more than two hundred thousand premature deaths across the EU member countries in 2020, so including this undesirable outcome during estimation is important to avoid biased results. We also extend the three-pillar model proposed by Boussemart et al. (2020) by accommodating multiple best-practice frontiers within a single pillar. Estimating our multi-frontier model using the nonparametric linear programming methodology and data on 28 OECD member countries for the period 2001 – 2019 yields the following conclusions.

First, we demonstrate that the overall well-being can be improved by an average of around 26 % when relatively similar weights are assigned to each performance pillar. By contrast, attributing disproportionate importance to any single pillar leaves relatively little room for improvement along the other two dimensions. Our results suggest that the most significant improvements can be realized in the areas of environmental performance and pollution-induced mortality, which is consistent with a general decline in environmental capital reported in UN's Inclusive Wealth (2018).

Second, we find that the European OECD member countries have fared generally better than the non-European OECD members, especially in terms of the basic health services coverage index and life expectancy, which can be attributed to the relatively strong safety nets characteristic of the European economies. However, the non-European OECD members have performed relatively well in the areas of fine particle pollution and life expectancy, which is in line with a relatively high intensity of per capita $PM_{2.5}$ damages observed in Europe during the last several decades (Inclusive Wealth, 2018).

Finally, our results suggest that while the purely economic performance gaps among the OECD countries have narrowed from more than 20 % to less than 15 % between 2001 and 2019, the average environmental performance and social well-being shortfalls have remained almost unchanged during the same period. Furthermore, performance gaps associated with particle pollution followed a generally increasing trajectory during the last decade, likely contributing to a rise in pollution-induced mortality. Some of the most significant temporal improvements can be observed in the area of basic health care coverage, where performance shortfalls dropped from almost 20 % in 2009 to 8 % in 2019. We believe these changes to be due to significantly wider health coverage offered by the U.S. Affordable Care Act starting from early 2010s.

These findings have important implications for policymakers looking to promote sustainable development trajectories targeting inclusive growth. They demonstrate that prioritizing any single performance dimension while ignoring the other well-being pillars is insufficient to achieve balanced growth. Moreover, they identify country-level differences in environmental inefficiency as the most important area of improvement when all three pillars are equally important.

Future work along this line of inquiry should focus on expanded specifications of the model that can accommodate additional performance outcomes. These can include desirable environmental outputs such as biodiversity indexes or socially undesirable outcomes unrelated to pollution, including corruption levels or measures of income inequality. Second, additional environmental disamenities such as deforestation levels or wastewater generation could be considered within the environmental sub-technology. Moreover, using multi-stage production frameworks to assess productivity change can offer additional insights into performance dynamics, because it accounts for technological improvements in addition to temporal changes in efficiency. Finally, looking beyond the developed nations' performance, the evolution in the inclusive growth of the developing countries can offer valuable insights into policies that can be used to promote their sustainable development.

CRediT authorship contribution statement

Yiran Niu: Data curation, Formal analysis, Investigation, Validation, Writing – original draft, Visualization. **Jean-Philippe Boussemart:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing, Methodology. **Zhiyang Shen:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Michael Vardanyan:** Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing.

References

Allen, K. (1999). Dea in the ecological context—An overview. *Data Envelopment Analysis In The Service Sector*, 203–235.

Ayres, R. U., & Kneese, A. V. (1969). Production, consumption, and externalities. *American Economic Review, 59*(3), 282–297.

Baumol, W. J. (1958). Activity analysis in one lesson. *American Economic Review, 68*, 665–672.

Boussemart, J. P., Leleu, H., Shen, Z., & Vardanyan, M. (2019). Decomposing banking performance into economic and credit risk efficiencies. *European Journal of Operational Research, 277*, 719–726.

Boussemart, J. P., Leleu, H., Shen, Z., & Valdmanis, V. (2020). Performance analysis for three pillars of sustainability. *Journal of Productivity Analysis, 53*(3), 305–320.

Callens, I., & Tyteca, D. (1999). Towards indicators of sustainable development for firms: A productive efficiency perspective. *Ecological Economics, 28*(1), 41–53.

Camanho, A. S., & Dyson, R. G. (2005). Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. *European Journal of Operational Research, 161*(2), 432–446.

Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory, 70*(2), 407–419.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.

Chen, & Ali, A. I (2004). DEA Malmquist productivity measure: New insights with an application to computer industry. *European Journal of Operational Research, 159*(1), 239–249.

Chen, J., & Chen, J. (2020). Measuring and improving eco-efficiency. *Environmental Modeling & Assessment, 25*(3), 373–395.

Chung, Y. H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management, 51*(3), 229–240.

Coelli, T., Lauwers, L., & Van Huylenbroeck, G. (2007). Environmental efficiency measurement and the materials balance condition. *Journal of Productivity Analysis, 28*(1), 3–12.

Coelli, T., Rao, D., & Battese, G. E. (1998). Efficiency measurement using data envelopment analysis (DEA). *An introduction to efficiency and productivity analysis* (pp. 133–160). Springer.

Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *The Lancet, 389*(10082), 1907–1918.

Considine, T. J., & Larson, D. F. (2006). The environment as a factor of production. *Journal of Environmental Economics and Management, 52*(3), 645–662.

Cook, W. D., & Zhu, J. (2007). Within-group common weights in DEA: An analysis of power plant efficiency. *European Journal of Operational Research, 178*(1), 207–216.

Cooper, W. W., Ruiz, J. L., & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research, 195*(2), 563–574.

Cropper, M. L., & Oates, W. E. (1992). Environmental economics: A survey. *Journal of Economic Literature, 30*(2), 675–740.

Dakpo, K. H., Jeanneaux, P., & Latruffe, L. (2016). Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework. *European Journal of Operational Research, 250*(2), 347–359.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, 273–292.

European Environment Agency (2022). *Air Quality in Europe 2022*. Report no. 05/2022, 10.2800/488115.

Färe, R., & Grosskopf, S. (1983). Measuring output efficiency. *European Journal of Operational Research, 13*(2), 173–179.

Färe, R., & Grosskopf, S. (2003). Nonparametric productivity analysis with undesirable outputs: Comment. *American Journal of Agricultural Economics, 85*(4), 1070–1074.

Färe, R., & Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation: Comment. *European Journal of Operational Research, 157*(1), 242–245.

Färe, R., & Grosskopf, S. (2006). *New directions: Efficiency and productivity, 3*. Springer Science & Business Media.

Färe, R., Grosskopf, S., Lovell, C. K., & Pasurka, C. (1989). Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach. *The Review of Economics and Statistics*, 90–98.

Färe, R., Grosskopf, S., Lundgren, T., Marklund, P.O., & Zhou, W. (2012). Productivity: Should we include bads? The Centre for Environmental and Resource Economics Working Paper 2012:13.

Färe, R., Grosskopf, S., & Pasurka, C. (1986). Effects on relative efficiency in electric power generation due to environmental controls. *Resources and Energy, 8*(2), 167–184.

Färe, R., Grosskopf, S., & Pasurka, C. A., Jr (2007). Environmental production functions and environmental directional distance functions. *Energy, 32*(7), 1055–1066.

Fare, R., Pasurka, C., & Vardanyan, M. (2017). On endogenizing direction vectors in parametric, directional distance function-based models. *European Journal of Operational Research, 262*(1), 361–369.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General), 120*(3), 253–281.

Førsund, F. R. (1998). *Pollution modelling and multiple-output production theory*. Department of Economics, University of Oslo.

Førsund, F. R. (1999). On the contribution of Ragnar Frisch to production theory. *Rivista Internazionale di Scienze Economiche e Commerciali, 46*, 1–34.

Førsund, F. R. (2009). Good Modelling of Bad Outputs: Pollution and Multiple-Output Production. *International Review of Environmental and Resource Economics, 3*(1), 1–38.

Førsund, F. R. (2018). Multi-equation modelling of desirable and undesirable outputs satisfying the materials balance. *Empirical Economics, 54*(1), 67–99.

Fukuyama, H., Hashimoto, A., & Weber, W. L. (2020). Environmental efficiency, energy efficiency and aggregate well-being of Japanese prefectures. *Journal of Cleaner Production, 271*, Article 122810.

Golany, B., Roll, Y., & Rybak, D. (1994). Measuring efficiency of power plants in Israel by data envelopment analysis. *IEEE Transactions on Engineering Management, 41*(3), 291–301.

Hailu, A., & Veeman, T. S. (2001). Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. *American Journal of Agricultural Economics, 83*(3), 605–616.

Hamilton, K. (1996). Pollution and pollution abatement in the national accounts. *The Review of Income and Wealth, 42*(1), 13–33.

Hampf, B. (2018). Measuring inefficiency in the presence of bad outputs: Does the disposability assumption matter? *Empirical Economics, 54*(1), 101–127.

Hampf, B., & Rødseth, K. L. (2019). Environmental efficiency measurement with heterogeneous input quality: A nonparametric analysis of US power plants. *Energy Economics, 81*, 610–625.

Hoang, V. N., & Coelli, T. (2011). Measurement of agricultural total factor productivity growth incorporating environmental factos: A nutrients balance approach. *Journal of Environmental Economics and Management, 62*(3), 462–474.

IHME, Global Burden of Disease Study (2019). Processed by our world in data.

Inclusive Wealth, Report. (2018). Inclusive wealth report 2018; measuring progress towards sustainability. *United Nations Environment Programme*. Taylor & Francis.

Intergovernmental Panel on Climate Change. (1990). *Climate change: The ipcc scientific assessment*. Cambridge University Press.

Koopmans, T. C. (1953). Activity analysis and its applications. *American Economic Review, 43*, 406–414.

Kamali, F. P., Meuwissen, M., de Boer, I., van Middelaar, C., Moreira, A., & Oude Lansink, A. (2017). Evaluation of the environmental, economic, and social performance of soybean farming systems in southern Brazil. *Journal of Cleaner Production, 142*(1), 385–394.

Leontief, W. (1970). Environmental repercussions and the economic structure: An input-output approach. *The Review of Economics and Statistics, 52*, 262–271.

Leontief, W. (1973). National income, economic structure, and environmental externalities. In Milton Moss (Ed.), *The measurement of economic and social performance* (pp. 565–576). National Bureau of Economic Research.

Levkoff, S. B. (2011). Decomposing NOx and SO_2 Electric Power Plant Emissions in a "By-production" Framework: A Nonparametric DEA Study. *Essays on efficiency measurement with environmental applications*. Riverside: University of California.

Li, M., Fu, Q., Singh, V. P., Liu, D., Li, T., & Zhou, Y. (2020). Managing agricultural water and land resources with tradeoff between economic, environmental, and social

considerations: A multi-objective non-linear optimization model under uncertainty. *Agricultural Systems, 178*, Article 102685.
Lovell, C. K., & Eeckaut, P. V. (1993). Frontier tales: DEA and FDH. *Mathematical modelling in economics* (pp. 446–457). Springer.
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 Journal of Operational Research, 87*(3), 507–518.
Mahlberg, B., & Sahoo, B. K. (2011). Radial and non-radial decompositions of Luenberger productivity indicator with an illustrative application. *International Journal of Production Economics, 131*(2), 721–726.
Mandal, S. K., & Madheswaran, S. (2010). Environmental efficiency of the Indian cement industry: An interstate analysis. *Energy Policy, 38*(2), 1108–1118.
Murty, S. (2015). On the properties of an emission-generating technology and its parametric representation. *Economic theory, 60*, 243–282.
Murty, S., & Russell, R. R. (2018). Modeling emission-generating technologies: reconciliation of axiomatic and by-production approaches. *Empirical Economics, 54*, 7–30.
Murty, S., Russell, R. R., & Levkoff, S. B. (2012). On modeling pollution-generating technologies. *Journal of Environmental Economics and Management, 64*(1), 117–135.
Nordhaus, W. D., & Tobin, J. (1972). Is Growth Obsolete?. In *Economic research: retrospect and prospect, 5* pp. 1–80) Chicago IL: Economic Growth, NBER Books.
Organisation for Economic Cooperation and Development. (2020). *How's life? 2020*. OECD.
Organisation for Economic Cooperation and Development. (2023). *Environment at a glance indicators*. OECD.
Pethig, R. (2003). The materials balance approach to pollution: Its origin, implications and acceptance. Leibniz Information Centre for Economics Working Paper Number 105-03.
Pethig, R. (2006). Non-linear production, abatement, pollution and materials balance reconsidered. *Journal of Environmental Economics and Management, 51*(2), 185–204.
Pham, M. D., & Zelenyuk, V. (2019). Weak disposability in nonparametric production analysis: A new taxonomy of reference technology sets. *European Journal of Operational Research, 274*(1), 186–198.
Rakshit, I., & Mandal, S. K. (2020). A global level analysis of environmental energy efficiency: An application of data envelopment analysis. *Energy Efficiency, 13*(5), 889–909.
Ray, S. C. (2004). *Data envelopment analysis: Theory and techniques for economics and operations research*. Cambridge University Press.
Ray, S. C. (2015). Nonparametric measures of scale efficiency and capacity utilization: An application to US manufacturing. *European Journal of Operational Research, 245*, 602–611.
Ray, S. C., Mukherjee, K., & Venkatesh, A. (2018). Nonparametric measures of efficiency in the presence of undesirable outputs: A by-production approach. *Empirical Economics, 54*(1), 31–65.
Ray, S. C., Walden, J., & Chen, L. (2021). Economic measures of capacity utilization: A nonparametric short run cost function analysis. *European Journal of Operational Research, 293*, 375–387.
Russell, R., & Murty, S. (2002). On modeling pollution-generating technologies. University of California, Riverside, Department of Economics, Working Paper Series, 02-14.
Shen, Z., Vardanyan, M., Baležentis, T., & Wang, J. (2024). Analyzing the tradeoff between the economic and environmental performance: the case of the chinese manufacturing sector. *IEEE Transactions on Engineering Management, 71*, 233–244.
Shen, Z., Baležentis, T., & Vardanyan, M. (2022a). Evaluating green productivity gains with the exponential by-production technology: An analysis of the Chinese industrial sector. *Environmental Modeling & Assessment, 27*, 759–770.
Shen, Z., Wu, H., Bai, K., & Hao, Y. (2022b). Integrating economic, environmental and societal performance within the productivity measurement. *Technological Forecasting and Social Change, 176*, Article 121463.
Shephard, R. W. (1953). *Cost and production functions*. Princeton, New Jersey: Princeton University Press.
Shephard, R. W., & Färe, R. (1974). The law of diminishing returns. *Production theory* (pp. 287–318). Springer.
Smith, P. (1990). Data envelopment analysis applied to financial statements. *Omega, 18* (2), 131–138.
Song, M., An, Q., Zhang, W., Wang, Z., & Wu, J. (2012). Environmental efficiency evaluation based on data envelopment analysis: A review. *Renewable and Sustainable Energy Reviews, 16*(7), 4465–4469.
Stiglitz, J.E., Sen, A., & Fitoussi, J.P. (2009). Report by the commission on the measurement of economic performance and social progress.
Stiglitz, J. E., Fitoussi, J. P., & Durand, M. (2018). *Beyond GDP: Measuring what counts for economic and social performance*. OECD Publishing.
Yang, H., & Pollitt, M. (2010). The necessity of distinguishing weak and strong disposability among undesirable outputs in DEA: Environmental performance of Chinese coal-fired power plants. *Energy Policy, 38*(8), 4440–4444.