Proceedings of ALGORITMY 2024
pp. 169–178

PATH-BASED DEA MODELS AND A SINGLE-STAGE APPROACH FOR FINDING AN EFFICIENT BENCHMARK*

JAKUB HRDINA †, MÁRIA TRNOVSKÁ , AND MARGARÉTA HALICKÁ.

Abstract. In data envelopment analysis (DEA) each model for efficiency evaluation can be formulated in two forms - the envelopment and the multiplier form that are in a primal-dual relationship. The general class of path-based DEA models, which also includes nonlinear convex models, is formulated in the envelopment form. In general, models of this class do not project onto the strongly efficient frontier, and hence a two-stage procedure is used to find a strongly efficient benchmark for the assessed unit. In this paper, we use the multiplier form of general path-based models to formulate a single-stage optimisation procedure for finding a strongly efficient benchmark. We illustrate the numerical tractability of the proposed approach on an environmental data set of 27 EU countries.

. **Key words.** Data envelopment analysis, path-based models, directional distance function model, strongly efficient benchmark

AMS subject classifications. 90C05, 90C25, 90C90

1. Introduction. Path-based models such as radial BCC input or output oriented models ([1]), directional distance function models ([2, 3]) and hyperbolic distance function model ([5]) search for benchmarks by specifying various parametric paths that run from the assessed unit to the boundary of the technology set. The point at which the path leaves the technology set is called the projection of the assessed units. Since the projection does not need to be strongly efficient, the so-called second stage is formulated for these models, which provides strongly efficient benchmarks for the assessed unit.

In the paper [11] a single-stage method was proposed to find a strongly efficient benchmark for the assessed unit in terms of radial input or output BCC models. The method combines the model of the first stage in its multiplier form with the traditional model of the second stage, which is formulated using a modified envelopment model. In the case of BCC models, the envelopment models are linear and their duals are always known. Thus, the resulting single-stage model of [11] led to a linear model. Path-based models are analysed in [9] using a general envelope scheme and are formulated in the form of convex programmes. The dual (multiplier) form of these models, derived in a recent work [12], makes it possible to modify the procedure from [11] also to convex path-based models.

The article is organised as follows. Section 2 presents all the necessary material to

*The research of the authors was supported by the APVV-20-0311 project of the Slovak Research and Development Agency and the VEGA 1/0611/21 grant administered jointly by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences.

†Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina F1, Bratislava, 84248, Slovakia (jakub.hrdina@fmph.uniba.sk, trnovska@fmph.uniba.sk, halicka@fmph.uniba.sk)

determine the main result of our article. Section 3 formulates the one-stage procedure and analyses its properties. Section 4 illustrates the numerical tractability of the proposed procedure on the example of the eco-data set of 27 EU countries. Section 5 concludes and discusses the advantages of the one-stage procedure.

2. Preliminaries. Consider a set of n observed decision making units DMU_j ($j = 1, \dots, n$), each consuming m inputs x_{ij} ($i = 1, \dots, m$) to produce s outputs y_{rj} ($r = 1, \dots, s$). For each $j = 1, \dots, n$, the data of inputs and outputs of DMU_j can be arranged into column vectors $\boldsymbol{x}_j = (x_{1j}, \dots, x_{mj})^\top \in \mathbb{R}^m$ of inputs and $\boldsymbol{y}_j = (y_{1j}, \dots, y_{sj})^\top \in \mathbb{R}^s$ of outputs. Finally, the input and output vectors of all DMUs form $m \times n$ input and $s \times n$ output matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, i.e., $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]$, respectively.

2.1. Technology set. Based on the given data set we define the technology set

$$\mathcal{T} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^m \times \mathbb{R}^s \mid \boldsymbol{X}\boldsymbol{\lambda} \leq \boldsymbol{x},\ \boldsymbol{Y}\boldsymbol{\lambda} \geq \boldsymbol{y},\ \boldsymbol{\lambda} \geq \boldsymbol{0},\ \boldsymbol{e}^\top \boldsymbol{\lambda} = 1 \right\}, \tag{2.1}$$

which corresponds to variable returns to scale (VRS). Note that the common nonnegativity of $(\boldsymbol{x}, \boldsymbol{y})$ is not imposed here. The input/output vectors of $\mathcal{T}$ will be called units. By $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ we denote the unit under evaluation.

It is said that unit $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ *dominates* unit $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ if and only if $\boldsymbol{x}_1 \leq \boldsymbol{x}_o$ and $\boldsymbol{y}_1 \geq \boldsymbol{y}_o$. Moreover, unit $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ is called *strongly efficient* if and only if there is no other unit in $\mathcal{T}$ dominating $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, i.e. if $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in \mathcal{T}$ dominates $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, then $(\boldsymbol{x}_1, \boldsymbol{y}_1) = (\boldsymbol{x}_o, \boldsymbol{y}_o)$. Obviously, a strongly efficient unit lies on the boundary of $\mathcal{T}$. The set of all strongly efficient units is called the efficient boundary of $\mathcal{T}$ and denoted $\partial^S \mathcal{T}$. The remaining part $\partial^W \mathcal{T}$ of the boundary of $\mathcal{T}$ consists of the so-called weakly efficient units.

2.2. Path-based model in the envelopment form. We follow up on the paper [9] and consider the general scheme (GS) model applied to $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ with directions $\boldsymbol{g}_o = (\boldsymbol{g}_o^x, \boldsymbol{g}_o^y) \gneqq 0$ that may depend on $(\boldsymbol{x}_o, \boldsymbol{y}_o)$:

$$(\mathrm{GS})_o \qquad \min \theta \tag{2.2a}$$

$$\boldsymbol{X}\boldsymbol{\lambda} \leq \boldsymbol{x}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^x, \tag{2.2b}$$

$$\boldsymbol{Y}\boldsymbol{\lambda} \geq \boldsymbol{y}_o + (\psi^y(\theta) - 1)\boldsymbol{g}_o^y, \tag{2.2c}$$

$$\boldsymbol{e}^\top \boldsymbol{\lambda} = 1, \quad \boldsymbol{\lambda} \geq \boldsymbol{0}. \tag{2.2d}$$

The real valued functions ψ, their domains (dom), and their images (im) satisfy the following assumptions: 1. $\mathrm{dom}(\psi^x), \mathrm{dom}(\psi^y)$ and $\mathrm{im}(\psi^x), \mathrm{im}(\psi^y)$ are either $(0, \infty)$ or $(-\infty, \infty)$; 2. ψ^x is smooth, concave, increasing and ψ^y is smooth, convex, decreasing; 3. $\psi^x(1) = \psi^y(1) = 1$; The right-hand sides of (2.2b) and (2.2c) denoted as

$$\boldsymbol{\phi}_o^x(\theta) := \boldsymbol{x}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^x \quad \text{and} \quad \boldsymbol{\phi}_o^y(\theta) := \boldsymbol{y}_o + (\psi^y(\theta) - 1)\boldsymbol{g}_o^y \tag{2.3}$$

define a continuous path $\boldsymbol{\phi}_o : \theta \mapsto (\boldsymbol{\phi}_o^x(\theta), \boldsymbol{\phi}_o^y(\theta))$ in the input-output space $\mathbb{R}^m \times \mathbb{R}^s$ parameterised by $\theta \in \mathcal{D} = \mathrm{dom}(\psi^x) \cap \mathrm{dom}(\psi^y)$. The path $\boldsymbol{\phi}_o$ passes through the point $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ at $\theta = 1$ and for decreasing values of θ it moves towards the boundary of $\mathcal{T}$. The path intersects the boundary $\partial \mathcal{T}$ at the *projection* point $\boldsymbol{\phi}_o(\theta_o^*)$, where $\theta_o^* \leq 1$ is the optimal value in $(GS)_o$ and is called the *efficiency score*.

The general scheme (2.2) includes the well-known models[1] such as the BCC input and output models [1], the hyperbolic distance function model (HDF) [5], where $\psi^x(\theta) = \theta, \psi^y(\theta) = 1/\theta$, the general directional distance function model (DDF) [2], where $\psi^x(\theta) = \theta, \psi^y(\theta) = 2 - \theta$, and also the so-called generalised distance function model (GDF) introduced by [4].

2.3. Strongly efficient benchmarks and the two stages procedure. Note that the projection belongs to the boundary of $\mathcal{T}$, but it is not necessarily strongly efficient. However, the programme (2.2) may have multiple optimal $\boldsymbol{\lambda}^*$, and the points $(\boldsymbol{X}\boldsymbol{\lambda}^*, \boldsymbol{Y}\boldsymbol{\lambda}^*)$ are called the *benchmarks* for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$. Among them, there are also strongly efficient benchmarks that can be computed using the standard two stages procedure described and analysed for path-based models in [9]. The first stage identifies the optimal value θ_o^* by solving (2.2), the second stage solves a modified envelopment programme that at the fixed optimal value of θ_o^* maximises the sum of slacks. In the context of the GS scheme the second stage programme is as follows:

$$\max\ \boldsymbol{e}^\top \boldsymbol{s}^x + \boldsymbol{e}^\top \boldsymbol{s}^y \tag{2.4a}$$
$$\text{s.t.}\quad \boldsymbol{X}\boldsymbol{\lambda} + \boldsymbol{s}^x \le \boldsymbol{x}_o + (\psi^x(\theta_o^*) - 1)\boldsymbol{g}_o^x, \tag{2.4b}$$
$$\boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{s}^y \ge \boldsymbol{y}_o + (\psi^y(\theta_o^*) - 1)\boldsymbol{g}_o^y, \tag{2.4c}$$
$$\boldsymbol{e}^\top \boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \ge \boldsymbol{0}, \boldsymbol{s}^x \ge \boldsymbol{0}, \boldsymbol{s}^y \ge \boldsymbol{0}. \tag{2.4d}$$

It was shown in [9, Theorem 4.7] that for each optimal solution $(\boldsymbol{\lambda}^*, \boldsymbol{s}^{x*}, \boldsymbol{s}^{y*})$ in (2.4), the unit $(\boldsymbol{X}\boldsymbol{\lambda}^*, \boldsymbol{Y}\boldsymbol{\lambda}^*)$ is a strongly efficient benchmark for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$.

2.4. The multiplier form of the path-based models. The general scheme for a path-based model is formulated in the envelopment form, where programme (2.2) presents a convex optimisation problem. In [12] the Lagrangian dual for (2.2) was derived, which represents the multiplier form of the path-based model. The dual can be written as follows:

$$(\mathrm{DGS})_o \qquad \max\ 1 - \boldsymbol{v}^\top \boldsymbol{x}_o + \boldsymbol{u}^\top \boldsymbol{y}_o - \sigma - F(\boldsymbol{u}, \boldsymbol{v}) \tag{2.5a}$$
$$\boldsymbol{Y}^\top \boldsymbol{u} - \boldsymbol{X}^\top \boldsymbol{v} \le \sigma \boldsymbol{e}, \tag{2.5b}$$
$$\boldsymbol{u}, \boldsymbol{v} \ge 0, \text{where} \tag{2.5c}$$
$$F(\boldsymbol{u}, \boldsymbol{v}) := 1 - \boldsymbol{v}^\top \boldsymbol{g}_o^x + \boldsymbol{u}^\top \boldsymbol{g}_o^y - \inf_{\theta \in \mathcal{D}} [\theta - \boldsymbol{v}^\top \boldsymbol{g}_o^x \psi^x(\theta) + \boldsymbol{u}^\top \boldsymbol{g}_o^y \psi^y(\theta)]. \tag{2.6}$$

Note that $F(\boldsymbol{u}, \boldsymbol{v})$ is a convex function, and its specific forms for praticular path-based models are provided in [12]. The primal-dual relationship between (2.4) and (2.5) is described in [12, Theorem 1] and its weak duality result will play an important role in the formation of the single-stage model.

3. Single-stage model for finding a strongly efficient benchmark.

In this section, we propose a single-stage model, which is a generalisation of the approach of [11] to the whole class of path-based models. Note that the approach of [11] was developed for linear models and was described as follows: change the fixed optimal value in the second phase programme to the variable and make it equal to the objective

[1]Some models must first be equivalently reformulated to fit the scheme.

function of the multiplier first stage programme with simultaneous incorporating the constraints of the dual. This procedure applied to the linear models led to a linear model.

3.1. Single-stage model in terms of path-based models. Now, we have to modify this procedure so that if the first stage programme is convex (not necessarily linear), then the single-stage programme is also convex. First, we change the fixed θ_o^* in the second stage programme (2.4) to the variable θ and change the equality constraints in (2.4) to inequalities. Then we make θ less than or equal to the objective function of the dual (multiplier) first stage programme (2.2). Finally, we incorporate the constrains of the dual (multiplier) first stage programme in the resulting programme, which is now stated as follows:

$$\max\ \boldsymbol{e}^\top \boldsymbol{s}^x + \boldsymbol{e}^\top \boldsymbol{s}^y \tag{3.1a}$$
$$\text{s.t.}\quad \boldsymbol{X}\boldsymbol{\lambda} + \boldsymbol{s}^x \le \boldsymbol{x}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^x, \tag{3.1b}$$
$$\boldsymbol{Y}\boldsymbol{\lambda} - \boldsymbol{s}^y \ge \boldsymbol{y}_o + (\psi^y(\theta) - 1)\boldsymbol{g}_o^y, \tag{3.1c}$$
$$\boldsymbol{e}^\top\boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \ge \boldsymbol{0}, \tag{3.1d}$$
$$1 - \boldsymbol{v}^\top \boldsymbol{x}_o + \boldsymbol{u}^\top \boldsymbol{y}_o - \sigma - F(\boldsymbol{u},\boldsymbol{v}) \ge \theta, \tag{3.1e}$$
$$\boldsymbol{Y}^\top \boldsymbol{u} - \boldsymbol{X}^\top \boldsymbol{v} \le \sigma \boldsymbol{e}, \tag{3.1f}$$
$$\boldsymbol{s}^x \ge \boldsymbol{0}, \boldsymbol{s}^y \ge \boldsymbol{0}, \boldsymbol{u} \ge 0, \boldsymbol{v} \ge 0, . \tag{3.1g}$$

Note that this model is a convex optimisation problem, since, by assumption, $\psi^x(\theta)$ is a concave function, $\psi^y(\theta)$ is a convex function, and $F(\boldsymbol{u},\boldsymbol{v})$ is convex.

It can be easily seen that if $(\boldsymbol{s}^x, \boldsymbol{s}^y, \theta, \boldsymbol{\lambda}, \boldsymbol{u}, \boldsymbol{v}, \sigma)$ is feasible for (3.1), then $(\theta, \boldsymbol{\lambda})$ is feasible for the primal (envelopment) model (2.2) and $(\boldsymbol{u}, \boldsymbol{v}, \sigma)$ is feasible for the dual (multiplier) model (2.5). The weak duality states $1 - \boldsymbol{v}^\top \boldsymbol{x}_o + \boldsymbol{u}^\top \boldsymbol{y}_o - \sigma - F(\boldsymbol{u},\boldsymbol{v}) \le \theta$. Therefore, the constraint (3.1e) enforces strong duality.

3.2. Properties of the single-stage approach. The structure of the programme (3.1) enables us to formulate a similar result to [11, Theorem 1].

THEOREM 3.1. *The combined vector* $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ *is an optimal solution of programme* (3.1) *if and only if* $(\boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ *is an optimal solution of the first-stage programme* (2.5)*,* θ^* *is the optimal value of the programme* (2.5) *and* $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \boldsymbol{\lambda}^*)$ *is an optimal solution of the second-stage programme* (2.4) *with* $\theta_o^* = \theta^*$*.*

Proof. Let $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ be optimal for (3.1). As stated above, (θ^*, λ^*) is feasible for (2.2) and $(\boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ is feasible for (2.5), and due to the weak duality between programmes (2.2) and (2.5), and constraint (3.1e) we have that

$$1 - (\boldsymbol{v}^*)^\top \boldsymbol{x}_o + (\boldsymbol{u}^*)^\top \boldsymbol{y}_o - \sigma^* - F(\boldsymbol{u}^*, \boldsymbol{v}^*) = \theta^*.$$

Furthermore, from the weak duality property we find that $(\boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ is optimal for (2.5), and $(\theta^*, \boldsymbol{\lambda}^*)$ is optimal for (2.2) with θ^* being the optimal value of both programmes. Note that from the optimality of $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$, it must hold that $\boldsymbol{X}\boldsymbol{\lambda}^* + \boldsymbol{s}^{x*} = \boldsymbol{\phi}_o^x(\theta^*)$ and $\boldsymbol{Y}\boldsymbol{\lambda}^* - \boldsymbol{s}^{y*} = \boldsymbol{\phi}_o^y(\theta^*)$, which implies that $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \boldsymbol{\lambda}^*)$ is feasible for (2.4) with θ_o^* fixed to θ^*. Furthermore, if $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \boldsymbol{\lambda}^*)$ were not optimal for (2.4) with θ_o^* fixed to θ^*, there would exist a vector $(\boldsymbol{s}'^x, \boldsymbol{s}'^y, \boldsymbol{\lambda}')$ feasible for (2.4) with $\boldsymbol{e}^\top \boldsymbol{s}'^x + \boldsymbol{e}^\top s'^y > \boldsymbol{e}^\top \boldsymbol{s}^{x*} + \boldsymbol{e}^\top \boldsymbol{s}^{y*}$. The same would hold for the combined vector $(\boldsymbol{s}'^x, \boldsymbol{s}'^y, \theta^*, \boldsymbol{\lambda}', \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$, which would contradict the optimality of $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$.

To show the converse, suppose that $(\boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ is optimal for (2.5) with an optimal value θ^*, and $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \boldsymbol{\lambda}^*)$ is optimal for (2.4). Apparently,

$$1 - (\boldsymbol{v}^*)^\top \boldsymbol{x}_o + (\boldsymbol{u}^*)^\top \boldsymbol{y}_o - \sigma^* - F(\boldsymbol{u}^*, \boldsymbol{v}^*) = \theta^*,$$

and thus $(\theta^*, \boldsymbol{\lambda}^*)$ is optimal for (2.2), and the value θ_o^* in (2.4) is fixed to θ^*. Hence, the combined vector $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ is feasible for (3.1). If this vector were not optimal for (3.1), there would exist a vector $(\boldsymbol{s}'^x, \boldsymbol{s}'^y, \theta', \boldsymbol{\lambda}', \boldsymbol{u}', \boldsymbol{v}', \sigma')$, without loss of generality we may assume that $\boldsymbol{X}\boldsymbol{\lambda}' + \boldsymbol{s}'^x = \boldsymbol{\phi}_o^x(\theta')$ and $\boldsymbol{Y}\boldsymbol{\lambda}' - \boldsymbol{s}'^y = \boldsymbol{\phi}_o^y(\theta')$, feasible for (3.1) with $\boldsymbol{e}^\top \boldsymbol{s}'^x + \boldsymbol{e}^\top s'^y > \boldsymbol{e}^\top \boldsymbol{s}^{x*} + \boldsymbol{e}^\top \boldsymbol{s}^{y*}$. Moreover, $(\boldsymbol{u}', \boldsymbol{v}', \sigma')$ would be feasible for (2.5) and $(\theta', \boldsymbol{\lambda}')$ would be feasible for (2.2). Owing to (3.1e) we have

$$1 - \boldsymbol{v}'^\top \boldsymbol{x}_o + \boldsymbol{u}'^\top \boldsymbol{y}_o - \sigma' - F(\boldsymbol{u}', \boldsymbol{v}') = \theta',$$

and from the weak duality property, $(\boldsymbol{u}', \boldsymbol{v}', \sigma')$ would be optimal for (2.5) and $(\theta', \boldsymbol{\lambda}')$ would be optimal for (2.2). Obviously, $\theta' = \theta^*$ and because $(\boldsymbol{s}'^x, \boldsymbol{s}'^y, \boldsymbol{\lambda}')$ is feasible for (2.4), there would be a contradiction to the optimality of $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \boldsymbol{\lambda}^*)$ in (2.4). ☐

The optimal solutions of (3.1) allow us to construct a strongly efficient benchmark for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ and find out whether or not θ^* overestimates the efficiency of $(\boldsymbol{x}_o, \boldsymbol{y}_o)$.

COROLLARY 3.2. *Let $(\boldsymbol{s}^{x*}, \boldsymbol{s}^{y*}, \theta^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^*, \boldsymbol{v}^*, \sigma^*)$ be any of the optimal solutions of program (3.1) for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$. Then $(\boldsymbol{X}\boldsymbol{\lambda}^*, \boldsymbol{Y}\boldsymbol{\lambda}^*)$ is a strongly efficient benchmark for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$. Moreover,*
(a) $\boldsymbol{s}^{x*} = 0$ and $\boldsymbol{s}^{y*} = 0$ if and only if $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ is projected on $\partial^S \mathcal{T}$;
(b) $\boldsymbol{s}^{x*} = 0$, $\boldsymbol{s}^{y*} = 0$, and $\theta^* = 1$ if and only if $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ is strongly efficient.

Note that the information we acquire from the single stage model (3.1) can be obtained alternatively by solving the multiplier model (2.5) instead of the envelopment model (2.2) of the traditional first stage and the second phase model (2.4). In this procedure, we do not get the projection $(\boldsymbol{\phi}_o^x(\theta^*), \boldsymbol{\phi}_o^y(\theta^*))$ directly, but it must be calculated by substituting the optimal value in (2.5) into the formulas (2.3).

4. Numerical application to the data set of 27 EU countries. To check the numerical tractability of the single-stage model (3.1) and to compare the time complexity of the single-stage approach with the two-step procedure, we have chosen a set of real macroeconomic data from 27 EU countries. Data consist of 3 inputs: labour, capital, and energy consumption; one desirable output: GDP; and 3 undesirable outputs: Greenhouse Gas (GHG) emissions, Particulate Matter with a diameter of 10 micrometers or less (PM10), and Waste. Data extracted from the World Development Indicators and the Eurostat database for the year 2020 are included in Table 4.1.

We conducted two studies on this data set. The first, which we call *Case Study I*, does not use the environmental data listed in Table 4.1 under the heading of undesirable outputs. In this study, we applied the DDF and HDF models with directions $\boldsymbol{g}_o^x = 0$ and $\boldsymbol{g}_o^y = \boldsymbol{y}_o$ to measure the output efficiency. With this selection of directions, the model measures an ability of a particular country to maximise the output given input levels consumed. The second study, *Case Study II*, uses the entire data set, including the three undesirable outputs. The aim in this study is to design a model so

TABLE 4.1
Data for 27 EU countries

	Inputs			Outputs	Undesirable Outputs		
	Labour force	Capital	Energy consumption	GDP	GHG	PM10	Waste
	# individuals	Mil. EUR	1000t oil equiv.	Mil. EUR	1000t CO2 equiv.	Tonne	Tonne
AT	4 638 300	95 140,2	24 872,9	380 888,5	73 910,8	26 788	68 906 034
BE	5 167 188	110 943,3	30 938,8	460 747,7	107 272,7	25 513	68 061 590
BG	3 311 854	11 750,2	9 499,7	61 607,7	48 044,9	44 525	116 387 350
HR	1 772 376	11 210,8	6 432,4	50 543,1	23 906,8	51 409	6 003 760
CY	651 740	4 666,5	1 525,9	22 086,6	8 579,9	1 765	2 221 809
CZ	5 375 292	57 290,8	23 766,2	215 805,4	113 719,5	37 402	38 486 186
DK	3 028 252	68 992,8	13 067,4	311 356,3	42 852,9	22 412	20 135 564
EE	700 236	8 174,8	2 725,5	27 430,0	11 407,1	12 664	16 170 358
FI	2 751 071	57 231,0	23 246,1	238 038,0	47 822,2	26 569	116 082 531
FR	30 379 167	539 458,0	127 822,1	2 317 832,0	392 328,7	247 518	310 373 987
DE	43 501 190	733 188,0	194 248,1	3 403 730,0	730 922,7	182 126	401 156 266
EL	4 643 796	19 939,9	14 470,3	165 015,7	75 464,5	56 342	28 358 897
HU	4 724 407	36 558,8	17 606,5	137 866,0	62 965,3	53 542	17 150 400
IE	2 432 234	157 893,7	10 854,5	375 249,6	59 056,3	30 287	16 192 033
IT	25 126 337	298 506,8	103 057,1	1 661 239,8	384 969,9	219 537	174 887 620
LV	988 585	6 752,1	3 798,2	30 109,5	10 496,2	26 298	2 852 792
LT	1 486 169	10 663,8	5 284,2	49 873,2	20 203,5	27 554	6 695 731
LU	322 041	10 771,4	3 270,8	64 524,3	9 029,9	1 672	9 215 222
MT	275 724	2 696,4	500,8	13 352,4	2 111,9	1 641	3 528 663
NL	9 502 134	172 937,0	41 872,5	796 530,0	164 787,1	27 652	125 138 771
PL	18 245 536	96 351,6	70 231,5	526 147,2	371 895,0	395 660	170 233 670
PT	5 166 305	38 509,8	15 156,6	200 518,9	58 149,9	56 956	16 601 514
RO	8 908 333	51 881,1	23 472,4	220 486,6	112 036,0	149 066	141 364 457
SK	2 712 322	18 210,2	9 611,1	93 444,1	37 233,8	24 024	12 775 926
SI	1 029 744	8 892,8	4 445,9	47 044,9	15 974,8	13 044	7 518 375
ES	22 838 137	228 532,0	72 322,6	1 119 010,0	272 244,4	211 698	105 624 359
SE	5 462 300	120 694,3	31 019,2	480 556,4	46 214,0	35 267	151 823 910

that it would measure the so-called output eco-efficiency, i.e. to measure an ability of a country to maximise the output and minimise the undesirable outputs given inputs levels consumed.

4.1. Models modifications for eco-efficiency measurement. Various procedures can be used when environmental factors such as smoke pollution or waste are required to be included in the efficiency measurement. Overviews of these procedures are provided, for example, in [8, 14, 13]. One approach allows handling these undesirable outputs (also called bads) as inputs (see e.g. [10]). For the *output* eco-efficiency measurement, it is necessary to choose a model that allows simultaneous increase of desirable outputs and reduction of undesirable outputs. Any path-based model can be used for these proposals when a zero-direction vector is selected for traditional inputs and positive-direction vectors are selected for undesirable outputs and traditional (desired) outputs.

In Case study II we will deal with the undesirable outputs as inputs to assess the output eco-efficiency using the DDF and HDF models. For these purposes, we need to introduce the notation, which takes into account this new type of factors included in the analysis, and describe the modifications of the models, which reflects the approach we have chosen for the Case study II.

We will use the following notation in the environmental assessment: p is the number of undesirable outputs; $\boldsymbol{b}_j \in \mathbb{R}^p$ is the vector of undesirable outputs for the

DMU_j, $j = 1, \ldots, n$; $\boldsymbol{B} \in \mathbb{R}^{p \times n}$ is the matrix whose columns are vectors of undesirable outputs of all DMUs. Moreover, we introduce the following p-dimensional vectors corresponding to the bads: $\boldsymbol{w}$ is the shadow weight; $\boldsymbol{s}^b$ is the slack variable; $\boldsymbol{g}^b$ is the direction vector. Since undesirable outputs are treated as inputs, we must choose the real function $\psi^b(\theta)$, which enters the definition of a path for bads, the same as for inputs.

With these notations, both the single-stage model (3.1) as well as the two steps models (2.5) or (2.2), and (2.4) presented in the previous sections can be modified and applied directly to path-based models to find strongly efficient benchmarks. The modification of the single-stage model is as follows:

$$
\begin{aligned}
& \max\ \boldsymbol{e}^\top \boldsymbol{s}^x + \boldsymbol{e}^\top \boldsymbol{s}^b + \boldsymbol{e}^\top \boldsymbol{s}^y && \text{(4.1a)}\\
& \text{s.t.}\quad \boldsymbol{X\lambda} + \boldsymbol{s}^x \leq \boldsymbol{x}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^x, && \text{(4.1b)}\\
& \qquad \boldsymbol{B\lambda} + \boldsymbol{s}^b \leq \boldsymbol{b}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^b, && \text{(4.1c)}\\
& \qquad \boldsymbol{Y\lambda} - \boldsymbol{s}^y \geq \boldsymbol{y}_o + (\psi^y(\theta) - 1)\boldsymbol{g}_o^y, && \text{(4.1d)}\\
& \qquad \boldsymbol{e}^\top \boldsymbol{\lambda} = 1,\ \boldsymbol{\lambda} \geq \boldsymbol{0}, && \text{(4.1e)}\\
& \qquad 1 - \boldsymbol{v}^\top \boldsymbol{x}_o - \boldsymbol{w}^\top \boldsymbol{b}_o + \boldsymbol{u}^\top \boldsymbol{y}_o - \sigma - F(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}) \geq \theta, && \text{(4.1f)}\\
& \qquad \boldsymbol{Y}^\top \boldsymbol{u} - \boldsymbol{X}^\top \boldsymbol{v} - \boldsymbol{B}^\top \boldsymbol{w} \leq \sigma \boldsymbol{e}, && \text{(4.1g)}\\
& \qquad \boldsymbol{s}^x \geq \boldsymbol{0}, \boldsymbol{s}^b \geq \boldsymbol{0}, \boldsymbol{s}^y \geq \boldsymbol{0}, \boldsymbol{u} \geq 0, \boldsymbol{w} \geq \boldsymbol{0}, \boldsymbol{v} \geq 0. && \text{(4.1h)}
\end{aligned}
$$

For our numerical experiments, we selected the DDF model, where $\psi^x(\theta) = \theta, \psi^y(\theta) = 2 - \theta$ and the HDF model, where $\psi^x(\theta) = \theta, \psi^y(\theta) = 1/\theta$. For these two models, explicit forms of the function $F(\boldsymbol{u}, \boldsymbol{v})$ were derived in [12, Table 5] and their modifications for the environmental assessment read:

$$
F(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \begin{cases} 0, & \text{if } \ \boldsymbol{v}^\top \boldsymbol{g}_o^x + \boldsymbol{u}^\top \boldsymbol{g}_o^y + \boldsymbol{w}^\top \boldsymbol{g}_o^b = 1 \\ +\infty, & \text{otherwise,} \end{cases}
$$

for the DDF model, and

$$
F(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \left(\sqrt{\boldsymbol{u}^\top \boldsymbol{g}_o^y} + \sqrt{1 - \boldsymbol{v}^\top \boldsymbol{g}_o^x - \boldsymbol{w}^\top \boldsymbol{g}_o^b}\right)^2,
$$

for the HDF model. The directional vectors are chosen as $\boldsymbol{g}_o^x = 0,\ \boldsymbol{g}_o^b = \boldsymbol{b}_o,\ \boldsymbol{g}_o^y = \boldsymbol{y}_o$.

To solve optimisation programmes, we applied the CVX Matlab-based modelling system [6, 7] on the normalised data.

4.2. Results and disscusion. In both case studies, the application of HDF to the single stage model numerically failed. It was probably caused by the structure of this nonlinear model, where the inequality (3.1e), or (4.1f) causes the failure of Slater condition, important for the interior point methods. Therefore, in the case of HDF (and probably also other nonlinear models), it is necessary to use the standard two-phase method, which is reliable even for the nonlinear models (see numerical results [12] and [9]). The application of the DDF model, the single stage method provided the same results as the two-stage methods in both case studies. The single stage method, however, reduced the computational time to 57% in average. Numerical results can be found in Table 4.2 for the Case study I and in Table 4.3 for the Case study II.

Table 4.2

Efficiency score and shares of the efficient units on the benchmark of the inefficient unit evaluated using the output oriented DDF model in Case Study I.

		Efficient countries							
	Score	DK	FR	DE	EL	IE	IT	LU	MT
AT	0.861	0.96	0	0.04	0	0	0	0	0
BE	0.953	0.88	0	0.05	0	0.07	0	0	0
BG	0.491	0	0	0	0.53	0	0	0	0.47
HR	0.384	0	0	0	0.34	0	0	0.33	0.33
CY	0.747	0	0	0	0.04	0	0	0.15	0.81
CZ	0.437	0	0	0	0.30	0	0.15	0.55	0
EE	0.181	0	0	0	0.04	0	0	0.59	0.37
FI	0.864	0.68	0	0	0	0	0.02	0.29	0
HU	0.272	0	0	0	0.62	0	0.07	0.31	0
LV	0.476	0	0	0	0.16	0	0	0.16	0.68
LT	0.504	0	0	0	0.25	0	0	0.44	0.30
NL	0.996	0.83	0	0.15	0	0	0.02	0	0
PL	0.906	0	0	0	0.73	0	0.27	0	0
PT	0.841	0	0	0	0.28	0	0.09	0.63	0
RO	0.506	0	0	0	0.74	0	0.12	0.14	0
SK	0.619	0	0	0	0.47	0	0.01	0.52	0
SI	0.730	0	0	0	0.17	0	0	0.41	0.42
ES	0.914	0.38	0	0.04	0	0	0.57	0	0
SE	0.938	0.82	0	0.06	0	0.12	0	0	0

Table 4.3

Efficiency score and shares of the efficient units on the benchmark of the inefficient unit evaluated using the eco-modified DDF model in the Study Case II. The efficient countries France, Germany and Spain do not contribute to the efficient benchmarks of the inefficient units.

		Efficient units except FR, DE and ES									
	Score	CY	DK	EL	IE	IT	LU	MT	NL	PT	SE
AT	0.896	0	0.56	0	0	0	0.07	0	0.21	0	0.16
BE	0.959	0	0.56	0	0.08	0	0.02	0	0.35	0	0
BG	0.619	0	0	0.34	0	0	0.40	0.26	0	0	0
HR	0.867	0.80	0	0.01	0	0	0	0	0	0.19	0
CZ	0.673	0	0.44	0	0	0.06	0.49	0	0	0	0
EE	0.491	0	0.02	0	0	0	0.30	0.67	0	0	0.02
FI	0.864	0	0.68	0	0	0.02	0.30	0	0	0	0
HU	0.739	0.35	0.32	0	0	0	0	0	0	0.32	0
LV	0.969	0.97	0.03	0	0	0	0	0	0	0	0
LT	0.866	0.71	0	0.03	0	0	0.12	0	0	0.14	0
PL	0.906	0	0	0.73	0	0.27	0	0	0	0	0
RO	0.637	0	0	0.16	0	0.14	0.70	0	0	0	0
SK	0.896	0.71	0	0.12	0	0.02	0	0	0	0.15	0
SI	0.9	0.55	0	0.06	0	0	0.36	0	0	0.03	0

The efficient countries are listed in the top row of both tables. The inefficient ones are listed in the vertical column together with their efficiency scores. Each row also includes the values of the intensity coefficients $\boldsymbol{\lambda}_j$ for each inefficient unit, which can be positive only for efficient countries in the two-stage as well as the single stage model.

From Table 4.2 we can see that there are 8 efficient countries in Case study I, of which only France does not contribute to the efficient benchmark of inefficient countries. On the other hand, Luxembourg contributes to the benchmark of nine inefficient countries. Eight countries have lower efficiency score than Slovakia, and the benchmark for Slovakia is made up of 3 countries, of which Luxembourg and Greece have the largest share.

From Table 4.3 we can see that in Case study II there were 13 eco-efficient countries. The increase in the number of efficient units in Case study II compared to Case study I is a natural consequence of the increase in the number of factors. In this case, three efficient countries, namely France, Germany, and Spain, do not contribute to the efficient benchmark of inefficient countries. On the contrary, Luxembourg is again the largest contributor. Nine EU countries have lower eco-efficiency than Slovakia, and the benchmark for Slovakia is composed of four countries, of which Cyprus has the largest share. Estonia had the lowest score in both case studies.

Finally, let us note that this application was intended solely to verify the numerical tractability of the proposed approach and to demonstrate the type of information that can be read from the results. It cannot be considered a real analysis of the environmental efficiency of EU countries.

5. Conclusion. We have found that the single stage method is not suitable for nonlinear models. For linear models, the single-stage method is computationally reliable and shows significant time savings compared to the two-stage method. In order to demonstrate the latter in general, additional numerical experiments must be conducted on different data sets.

Acknowledgments. The authors thank Jana Szolgayová for her assistance in the sourcing and processing of relevant data for this study.

REFERENCES

[1] R. D. Banker, A. Charnes, and W. W. Cooper, *Some models for estimating technical and scale inefficiencies in data envelopment analysis*, Management Science, 30 (1984), pp. 1078–1092.

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

[3] ———, *Profit, directional distance functions, and Nerlovian efficiency*, Journal of Optimization Theory and Applications, 98 (1998), pp. 351–364.

[4] J.-P. Chavas and T. L. Cox, *A generalized distance function and the analysis of production efficiency*, Southern Economic Journal, 66 (1999), pp. 294–318.

[5] R. Färe, S. Grosskopf, and C. A. K. Lovell, *The Measurement of Efficiency of Production*, Springer Netherlands, 1985.

[6] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in Recent Advances in Learning and Control, V. Blondel, S. Boyd, and H. Kimura, eds., Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110. http://stanford.edu/~boyd/graph_dcp.html.

[7] M. Grant, S. Boyd, and Y. Ye, *Cvx: Matlab software for disciplined convex programming, version 2.0 beta*, 2013.

[8] M. Halická and M. Trnovská, *Negative features of hyperbolic and directional distance models for technologies with undesirable outputs*, Central European Journal of Operations Research (CEJOR), 26 (2018), pp. 887–907.

[9] M. Halická, M. Trnovská, and A. Černý, *A unified approach to radial, hyperbolic, and directional efficiency measurement in data envelopment analysis*, European Journal of Operational Research, 312 (2024), pp. 298–314.

[10] P. J. Korhonen and M. Luptacik, *Eco-efficiency analysis of power plants: An extension of data envelopment analysis*, European journal of operational research, 154 (2004), pp. 437–446.

[11] G. Papaioannou and V. V. Podinovski, *A single-stage optimization procedure for data envelopment analysis*, European Journal of Operational Research, 313 (2024), pp. 1119–1128.

[12] M. Trnovská, M. Halická, and J. Hrdina, *Path-based dea models in multiplier form and returns-to-scale analysis*, (2024). https://www.researchgate.net/publication/

377273268_Path-based_DEA_models_in_multiplier_form_and_returns-to-scale_analysis.

[13] H. Zhou, Y. Yang, Y. Chen, and J. Zhu, *Data envelopment analysis application in sustainability: The origins, development and future directions*, European journal of operational research, 264 (2018), pp. 1–16.

[14] P. Zhou, B. W. Ang, and K. L. Poh, *A survey of data envelopment analysis in energy and environmental studies*, European journal of operational research, 189 (2008), pp. 1–18.