Proceedings of ALGORITMY 2024
pp. 159–168

PATH-BASED DEA MODELS WITH DIRECTIONS DEFINED USING THE ANTI-IDEAL POINT *

MÁRIA TRNOVSKÁ†, MARGARÉTA HALICKÁ† , AND JANA SZOLGAYOVÁ†

Abstract. Data envelopment analysis (DEA) is a non-parametric technique for relative efficiency evaluation. It formulates models in the form of optimisation problems where the objective function can be interpreted as an efficiency measure and the optimal value is the efficiency score. Path-based DEA models represent a subclass of models where the efficiency score is found by following a parametric path running towards the boundary of the technology set. In this paper, we focus on models where the parametric path is characterised by a direction vector defined using the anti-ideal point. We show that these models possess several desirable properties and are applicable even in the presence of negative data. The results are illustrated with both a simple example and numerical experiments on real environmental data sets from 27 European countries.

Key words. Data envelopment analysis, path-based models, Directional distance function model

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

1. Introduction. This article follows on from two recently published papers ([11, 14]) analysing the properties of path-based models over general data. In [11], the models are formulated through a general scheme that includes several well-known models, such as input or output radial models, directional distance function (DDF) models, and the hyperbolic distance function (HDF) model. The paper provides theoretical tools allowing us to analyse the properties of individual models. The properties of the path-based model are determined by real-valued functions and direction vectors that together define the paths. The paper [14] analyses super-efficiency scores of path-based models, which may not be well defined for standard directions. It proposes a new direction vector defined using the anti-ideal point (AIP-direction) and shows that super-efficiency models defined using this direction are not only well defined but also possess several desirable properties. In this contribution, we complete the analysis and examine the path-based efficiency models defined using the AIP-direction. Here, in particular, analyses of the well-definedness, boundedness, and monotonicity of scores require special attention. We also illustrate the results with both a simple example and numerical experiments on real environmental data sets from 27 European countries.

The article is organised as follows. Section 2 presents all the necessary materials to determine the main result of our article. Section 3 analyses the individual properties of the super-efficient score with the AIP-directions and compares them with the

*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 (trnovska@fmph.uniba.sk, halicka@fmph.uniba.sk, szolgayova@fmph.uniba.sk)

properties of other directions. Section 4 illustrates the results with a simple example and a real environmental data set of 27 EU countries. Section 5 concludes and discusses the advantages of the AIP-directions.

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

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

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

which corresponds to variable returns to scale (VRS). Note that the common non-negativity assumption of $(\boldsymbol{x}, \boldsymbol{y})$ is not imposed here. The symbol $\boldsymbol{e}$ denotes a vector of ones. The input-output vectors $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^m \times \mathbb{R}^s$ will be called units or points. By $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ we denote the unit under evaluation.

A unit $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ *dominates* the unit $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ if and only if $\boldsymbol{x}_1 \leq \boldsymbol{x}_o$ and $\boldsymbol{y}_1 \geq \boldsymbol{y}_o$. A unit $(\boldsymbol{x}_1, \boldsymbol{y}_1)$ *strictly dominates* the unit $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ if and only if $\boldsymbol{x}_1 < \boldsymbol{x}_o$ and $\boldsymbol{y}_1 > \boldsymbol{y}_o$. Moreover, a unit $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ is called *strongly efficient* if and only if there is no other unit in $\mathcal{T}$ that dominates $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, 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 $\mathcal{T}$. The remaining part of the boundary of $\mathcal{T}$ consists of the so-called *weakly efficient* units.

2.2. Path-based models. For $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ and a direction $\boldsymbol{g}_o = (\boldsymbol{g}_o^x, \boldsymbol{g}_o^y) \gneqq 0$ that may depend on $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, we define $\boldsymbol{\phi}_o : \theta \mapsto (\boldsymbol{\phi}_o^x(\theta), \boldsymbol{\phi}_o^y(\theta))$ by prescription:

$$\boldsymbol{\phi}_o^x(\theta) := \boldsymbol{x}_o + (\psi^x(\theta) - 1)\boldsymbol{g}_o^x \quad \text{and} \quad \boldsymbol{\phi}_o^y(\theta) := \boldsymbol{y}_o + (\psi^y(\theta) - 1)\boldsymbol{g}_o^y. \tag{2.2}$$

Here, the real functions ψ with 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. $\psi^x(1) = \psi^y(1) = 1$. Note that (2.2) defines a continuous path in the input-output space $\mathbb{R}^m \times \mathbb{R}^s$ parametrised by θ, where $\theta \in \mathcal{D} = \mathrm{dom}(\psi^x) \cap \mathrm{dom}(\psi^y)$.

By 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 $(\boldsymbol{x}_o, \boldsymbol{y}_o)$, we understand

$$(\mathrm{GS})_o \qquad \min\{\theta \quad | \quad (\boldsymbol{\phi}_o^x(\theta), \boldsymbol{\phi}_o^y(\theta)) \in \mathcal{T}\}. \tag{2.3}$$

According to [11], assumptions 1.-3. imposed on the functions ψ^x and ψ^y, the assumption on directions $\boldsymbol{g}_o = (\boldsymbol{g}_o^x, \boldsymbol{g}_o^y) \gneqq 0$, and $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ guarantee the well-definedness of the programme $(GS)_o$ as well as other useful properties of the path $\boldsymbol{\phi}_o$: the path $\boldsymbol{\phi}_o$ passes through the point $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ at $\theta = 1$ (i.e., $\boldsymbol{\phi}_o(1) = (\boldsymbol{x}_o, \boldsymbol{y}_o)$), and for decreasing values of θ it moves toward the boundary of T gradually passing through points that dominate one another. Finally, the path leaves $\mathcal{T}$ at some $\boldsymbol{\phi}_o(\theta_o^*) \in \partial\mathcal{T}$, where $\theta_o^* \le 1$, and θ_o^* is the optimal value in $(GS)_o$. The optimal value θ_o^* is called the *efficiency score* for $(\boldsymbol{x}_o, \boldsymbol{y}_o)$. The point $(\boldsymbol{\phi}_o^x(\theta_o^*), \boldsymbol{\phi}_o^y(\theta_o^*))$ on the path $\boldsymbol{\phi}_o$ is called the *projection* of $(\boldsymbol{x}_o, \boldsymbol{y}_o)$ in the GS model.

It is easy to see that the BCC input and output models [2], the general hyperbolic distance function model (HDF-g) [11, 6] as well as the general directional distance measure model (DDF-g) [3] can be equivalently rewritten in the form of the general scheme (2.3). The scheme also includes the so-called generalised distance function model (GDF) introduced by [5]. The corresponding parameterizations are shown in Table 2.1.

TABLE 2.1
Parameterization of the standard path-based models.

Model	$\boldsymbol{\phi}_o^x(\theta)$	$\boldsymbol{\phi}_o^y(\theta)$
BCC-I [2]	$\boldsymbol{x}_o + (\theta - 1)\boldsymbol{x}_o$	$\boldsymbol{y}_o$
BCC-O [2]	$\boldsymbol{x}_o$	$\boldsymbol{y}_o + (\frac{1}{\theta} - 1)\boldsymbol{y}_o$
DDF-g [3]	$\boldsymbol{x}_o + (\theta - 1)\boldsymbol{g}_o^y$	$\boldsymbol{y}_o + (2 - \theta - 1)\boldsymbol{g}_o^y$
HDF-g [11, 6]	$\boldsymbol{x}_o + (\theta - 1)\boldsymbol{g}_o^x$	$\boldsymbol{y}_o + (\frac{1}{\theta} - 1)\boldsymbol{g}_o^y$
GDF [5] $p \in [0, 1]$	$\boldsymbol{x}_o + (\theta^{1-p} - 1)\boldsymbol{x}_o$	$\boldsymbol{y}_o + (\theta^{-p} - 1)\boldsymbol{y}_o$

2.3. Super-efficiency measurement. The scheme (2.3) is well defined for the evaluation of unit $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$, providing the efficiency score $\theta_o^* \le 1$. This scheme can also be applied to units $(\boldsymbol{x}_o, \boldsymbol{y}_o) \notin \mathcal{T}$ to assess its super-efficiency. Note that in this case, the programme (2.3) may not be well defined (it may be infeasible) and the well-definedness of particular path-based super-efficiency models must be assessed individually. For this purpose, Theorem 1 of [14] can be used, which formulates the necessary and sufficient conditions of feasibility. If a super-efficiency path-based model is feasible, then the corresponding super-efficiency score satisfies $\theta_o^* \ge 1$.

2.4. Directions. Although the data in $\boldsymbol{X}$ and $\boldsymbol{Y}$ may contain negative values, directions $\boldsymbol{g}_o = (\boldsymbol{g}_o^x, \boldsymbol{g}_o^y)$ must be nonnegative and nonzero even though they may depend on the assessed unit $(\boldsymbol{x}_o, \boldsymbol{y}_o)$. The directions considered in this study are listed in Table 2.2. For a description of the input part of the direction, we have used virtual input vectors $\boldsymbol{x}^{\min}$ and $\boldsymbol{x}^{\max}$, defined for $i = 1, \dots m$ as $x_i^{\min} = \min_j x_{ij}$ and $x_i^{\max} = \max_j x_{ij}$. The notation $\boldsymbol{y}^{\min}$ and $\boldsymbol{y}^{\max}$ is introduced analogously for the output vectors. The points $(\boldsymbol{x}^{\min}, \boldsymbol{y}^{\max})$ and $(\boldsymbol{x}^{\max}, \boldsymbol{y}^{\min})$ are called *ideal* (IP) and *anti-ideal* (AIP) points, respectively. In Table 2.2 we can see several possible directions choices. The direction (G3) is based on the ideal and anti-ideal points, and the direction (G2) on the current unit and the ideal point. The absolute values in the (G2) directions are due to super-eficiency purposes. The direction (G6) (also called the AIP direction) designed primarily for the super-efficiency needs in [14] is based on the current unit and anti-ideal point.

Next, we analyse the properties of the score for the path-based models with direction (G6) and compare it with the ones with the other directions (G1)–(G5). We will focus on the well-definedness, unit invariance, translation invariance, boundedness, and monotonicity of the efficiency score.

TABLE 2.2

The directions (G1)–(G5) are commonly encountered choices of directions g_o. The direction (G6) is a relatively new one. Here $\beta = 1$ if $\operatorname{im}(\psi^y) = (-\infty, \infty)$ (DDF-g model) and $\beta \geq 1$ if $\operatorname{im}(\psi^y) = (0, \infty)$ (HDF-g model).

Notation	$g^x_{io}, i \in \{1, \dots, m\}$	$g^y_{ro}, r \in \{1, \dots, s\}$	Reference
(G1)	$\lvert x_{io}\rvert$	$\lvert y_{ro}\rvert$	[4], [12], [11]
(G2)	$\lvert x_{io} - x_i^{\min}\rvert$	$\lvert y_r^{\max} - y_{ro}\rvert$	[13], [11]
(G3)	$x_i^{\max} - x_i^{\min}$	$y_r^{\max} - y_r^{\min}$	[13]
(G4)	$\frac{1}{n}\sum_j \lvert x_{ij}\rvert$	$\frac{1}{n}\sum_j \lvert x_{ij}\rvert$	[1], [11]
(G5)	1	1	[4]
(G6)	$\max\{x_i^{\max} - x_{io}, 0\}$	$\max\{\beta(y_{ro} - y_r^{\min}), 0\}$	[14]

3. Analysis of models with AIP directions (G6). The general feature of the path-based model is that efficiency scores equal to 1 show weak efficiency, but not necessarily strong efficiency of the assessed unit; in addition, projection points may not be strongly efficient, indicating that efficiency scores are overestimated. This means that the so-called properties of *indication* and *strong efficiency of projections* are not satisfied, see [11]. It can be seen from [10] that (under mild assumptions) the strong efficiency property is equivalent to *strict monotonicity* of the efficiency score. Therefore, even the property of strict monotonicity is not satisfied for this class of models, which means that improving the input or output may not lead to a higher efficiency score. Other desirable properties of models with the AIP directions (G6) must be examined individually.

3.1. Well-definedness. The AIP directions were developed in the work [14] for the purpose of super-efficiency assessment. In the super-efficiency models, the assessed unit is outside the interior of the technology set, the AIP directions are non-zero, and the corresponding super-efficiency models are well defined. On the other hand, in the efficiency assessment where $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$, the AIP directions (G6) may be zero for certain virtual units from $\mathcal{T}$ and so the corresponding efficiency models are not well-defined. For this reason, we will analyse the models only for units from the reduced technology set:

$$\mathcal{T}^+ = \{(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T} \mid \boldsymbol{x}_o \leq \boldsymbol{x}^{\max}, \boldsymbol{y}_o \geq \boldsymbol{y}^{\min}, (\boldsymbol{x}_o, \boldsymbol{y}_o) \neq (\boldsymbol{x}^{\max}, \boldsymbol{y}^{\min})\} \tag{3.1}$$

For points $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}^+$, the formula defining (G6) directions simplifies to

$$(G6): \qquad g^x_o = \boldsymbol{x}^{\max} - \boldsymbol{x}_o, \text{ and } g^y_o = \beta(\boldsymbol{y}_o - \boldsymbol{y}^{\min}), \text{ where } \beta \geq 1. \tag{3.2}$$

Note that the restriction to the set $\mathcal{T}^+$ will not affect the applicability of these directions in practice, since the efficiencies are evaluated only for units that generate the technology set. Except for the highly improbable case $(\boldsymbol{x}_o, \boldsymbol{y}_o) = (\boldsymbol{x}^{\max}, \boldsymbol{y}^{\min})$, such units belong to $\mathcal{T}^+$ and therefore the efficiency scores are well-defined. For the case $(\boldsymbol{x}_o, \boldsymbol{y}_o) = (\boldsymbol{x}^{\max}, \boldsymbol{y}^{\min})$ we define $\theta^*_o = -\infty$ provided $\operatorname{im}(\psi^y) = \mathbb{R}$, or $\theta^*_o = 0$ provided $\operatorname{im}(\psi^y) = (0, \infty)$.

3.2. Unit and translation invariance. The path-based models with (G6) directions satisfy the sufficient conditions for unit- and translation- invariance provided in [11, Theorem 7] and [11, Theorem 8], respectively, and hence are unit- and translation-invariant.

3.3. Boundedness. If $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$, then automatically $\theta_o^* \leq 1$ (since $\theta = 1$ is a feasible solution). On the other hand, satisfying the property $\theta_o^* \geq 0$ depends on the domains of ψ^x and ψ^y. With $\mathrm{dom}(\psi^x) = (0, \infty)$ or $\mathrm{dom}(\psi^y) = (0, \infty)$, which only occur for nonlinear ψ^x and ψ^y, one has $\theta_o^* > 0$ for any choice of g_o. However, with linear ψ one has $\mathrm{dom}(\psi) = \mathbb{R}$, and hence this property may fail. In fact, linear models with (G6) directions generally do not meet the sufficient condition for boundedness provided in [11, Theorem 6]. On the other hand, [11, Theorem 6] allows us to specify the units for which the score is not negative.

PROPOSITION 3.1. *Consider a DDF-g model with directions (G6) and denote* $(\boldsymbol{x}^c, \boldsymbol{y}^c) = \frac{1}{2}(\boldsymbol{x}^{\min}, \boldsymbol{y}^{\max}) + \frac{1}{2}(\boldsymbol{x}^{\max}, \boldsymbol{y}^{\min})$. *Then for all units* $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}_+$ *that are not strictly dominated by* $(\boldsymbol{x}^c, \boldsymbol{y}^c)$, *it holds* $\theta_o^* \geq 0$.

Proof. Assume that $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}_+$ is not strictly dominated by $(\boldsymbol{x}^c, \boldsymbol{y}^c)$. That means that either there exists i such that $\boldsymbol{x}_{io} \leq \boldsymbol{x}_i^c < \boldsymbol{x}_i^{\max}$ or there exists r such that $\boldsymbol{y}_{ro} \geq \boldsymbol{y}_r^c > \boldsymbol{y}_r^{\min}$. The first condition is equivalent to the existence of i such that $\boldsymbol{g}_{io}^x = \boldsymbol{x}_i^{\max} - \boldsymbol{x}_{io} > 0$ and $\boldsymbol{x}_i^{\min} + (\boldsymbol{x}_i^{\max} - \boldsymbol{x}_{io}) - \boldsymbol{x}_{io} \geq 0$ and the second condition is equivalent to the existence of r such that $\boldsymbol{g}_{ro}^y = \boldsymbol{y}_{ro} - \boldsymbol{y}_r^{\min} > 0$ and $\boldsymbol{y}_{ro} + (\boldsymbol{y}_{ro} - \boldsymbol{y}_r^{\min}) - \boldsymbol{y}_r^M \geq 0$. The two conditions correspond to the ones in [11, Theorem 6], which guarantee the property $\theta_o^* \geq 0$. ∎

3.4. Monotonicity. The monotonicity property of a DEA model states that the efficiency score of the dominated unit is not greater than the efficiency score of the dominanting unit.

PROPOSITION 3.2. *The path-based models with (G6) directions satisfy the property of monotonicity.*

Proof. According to [11, Theorem 9] it suffices to prove that for any $(\boldsymbol{x}_o, \boldsymbol{y}_o) \in \mathcal{T}$ the corresponding paths $\boldsymbol{\phi}_o^x(\theta)$ and $\boldsymbol{\phi}_o^y(\theta)$ are component-wise increasing in $\boldsymbol{x}_o$ and $\boldsymbol{y}_o$ at any fixed $\theta \leq 1$. It is straightforward to prove that these conditions are met. Actually, for $\boldsymbol{\phi}_o^x$ we obtain the following:

$$\boldsymbol{\phi}_o^x(\theta) = \boldsymbol{x}_o + (\psi^x(\theta) - 1)(\boldsymbol{x}^{\max} - x_o) = (2 - \psi^x(\theta))\boldsymbol{x}_o + k_1, \tag{3.3}$$

where k_1 involves terms independent of $\boldsymbol{x}_o$. Since ψ^x is increasing and $\psi^x(1) = 1$ we obtain $(2 - \psi^x(\theta)) > 0$ for all $\theta \leq 1$ and therefore $\boldsymbol{\phi}_o^x$ increases at each component. Similarly, for $\boldsymbol{\phi}_o^y$ we obtain the following:

$$\boldsymbol{\phi}_o^y(\theta) = \boldsymbol{y}_o + (\psi^y(\theta) - 1)\beta(y_o - y^{\min}) = (1 - \beta + \psi^y(\theta)\beta)\boldsymbol{y}_o + k_2, \tag{3.4}$$

where k_2 involves terms independent of $\boldsymbol{y}_o$. Since ψ^y is decreasing and $\psi^y(1) = 1$ we obtain $(1 - \beta + \psi^y(\theta)\beta) > 0$ for $\theta \leq 1$.∎

3.5. Super-efficiency. The super-efficiency of path-based models with (G6) directions was analysed in detail in [14]. It was shown that the super-efficiency model is well defined for all $(\boldsymbol{x}_o, \boldsymbol{y}_o) \notin \mathcal{T}$, that the score is bounded above by 2, and has the properties of unit and translation invariance and monotonicity. These properties were compared to the properties of the super-efficiency DDF-g and HDF-g models with directions (G1)–(G5) in [14, Table 3] and the models with (G6) directions were the only ones that met all the above-mentioned characteristics.

3.6. Comparison with the other directions. The results of our analysis for (G6) directions applied to the DDF-g and HDF-g models are compared with the properties of these models with (G1)–(G5) directions in Table 3.1. Here the results for the (G1)–(G5) directions are obtained from [11] and [14]. We can see that the efficiency score for the non-linear HDF-g model has the best properties precisely with (G6) directions - it satisfies all the observed properties. In the case of linear models, the disadvantage of the (G6) directions is that, unlike the (G1)–(G3) directions, they do not provide the scores bounded from below. However, the advantage is that they have better super-efficiency properties.

TABLE 3.1

Properties of the DDF-g model and the HDF-g model (in brackets) with respect to different choices of directions (G1)–(G6). ✓* *– the property is satisfied for positive directions;* ✓** *– the property is satisfied for positive data but not for general data;*

Direction/Property	(G1)	(G2)	(G3)	(G4)	(G5)	(G6)
$\theta_o^* \in [0, 1]$	✓**(✓)	✓(✓)	✓(✓)	×(✓)	×(✓)	×(✓)
unit invariance	✓(✓)	✓(✓)	✓(✓)	✓(✓)	×(×)	✓(✓)
translation invariance	×(×)	✓(✓)	✓(✓)	×(×)	✓(✓)	✓(✓)
monotonicity	✓**(✓)	✓(✓)	✓(✓)	✓(✓)	✓(✓)	✓(✓)
WD of super-efficiency	✓*(✓**)	✓*(×)	✓*(×)	✓*(×)	✓(×)	✓(✓)

4. Illustrative and numerical examples.

4.1. Illustrative example. Consider a single input and single output VRS technology generated by 7 units: A = (2, 2), B = (3, 6), C = (7, 10), D = (4, 1), E=(8, 3), F=(9, 5) and G=(4, 4) shown in Figure 4.1. Units A, B, C are strongly efficient, and the projections of inefficient units D, E, F, G to the frontier of $\mathcal{T}$ in the DDF-g and HDF-g models with AIP directions are outlined by the dot-dashed red lines. Inefficient units D and F have one component of the AIP directions zero, and their weakly efficient projections do not differ in the DDF-g and HDF-g models. The efficiency scores shown in the following table are in agreement with our theoretical results provided in Proposition 3.1: The units E and F are strictly dominated by unit (x^c, y^c) and their DDF-g efficiency scores are negative.

DMU_o	A	B	C	D	E	F	G
DDF-g efficiency (AIP directions)	1	1	1	0.6	-1.66	$-0,25$	0.74
HDF-g efficiency (AIP directions)	1	1	1	0.6	0.22	$0,44$	0.75

4.2. Numerical example. For our numerical experiments, we have chosen a data set of 27 EU countries provided in Table 4.1. The data contain three inputs: labour, capital, and energy consumption; one desirable output: GDP; and two unde-

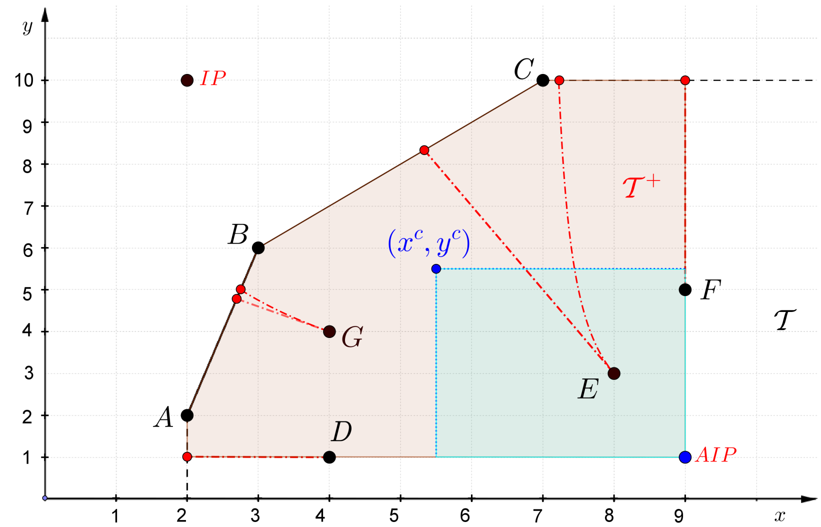

FIGURE 4.1. *The projections of inefficient units D, E, F and G to the frontier of $\mathcal{T}$ by DDF-g and HDF-g models with (G6) directions in Example 4.1.*

sirable outputs: CO2 including LULUCF and PM2.5[1]. Note that the first undesirable output also contains one negative value in our data set.

The data are extracted from the World Development Indicators and the Eurostat database, for the year 2021. We applied the DDF-g and HDF-g models with AIP directions (with $\beta = 1$) both to the data set without undesirable outputs and to the entire data set. In the environmental case study, we used the approach for modelling undesirable outputs (for an overview of approaches, see, e.g. [9, 16, 15]), which allowed us to deal with undesirable outputs as inputs. To solve optimisation programmes, we used the CVX Matlab-based modelling system [7, 8] on normalised data. The numerical results are given in Table 4.2.

Observe that the results in the DDF-g and HDM-g models (rounded to three decimal places) are almost the same. However, the DDF-g model is a linearisation of the HDF-g model around the evaluated unit, and thus the similarity of the results is a consequence of the distribution of inefficient units near the efficient boundary. Ten countries were identified as efficient in the standard models, 13 in the environmental models, and the super-efficiencies [2] were also calculated for efficient countries in all models. Germany showed the highest values in all models. An increase in the number of efficient countries and an increase in the score values in the environmental models compared to the standard models is a natural consequence of the increase in the number of input-output factors from 4 to 6. In addition, all countries have quite high

[1]The indicator CO2 including LULUCF refers to carbon dioxide emissions including Land Use, Land-Use Change, and Forestry. It can take negative values when there is a net removal of carbon dioxide from the atmosphere due to the land-use changes and forestry activities. The PM2.5 indicator refers to the concentration of particulate matter with a diameter of 2.5 micrometers or smaller in the air.

[2]The super-efficiency of an efficient unit was calculated with respect the modified technology set, where the efficient unit was removed from the set of units generating $\mathcal{T}$.

TABLE 4.1
Input and Output variables data for 27 EU countries

	Inputs			Outputs		
	Labour force	Capital	Energy consumption	GDP	CO2 incl. LULUCF	PM2.5
DMU	(# individuals)	(Million EUR)	(1000 tonnes of oil equiv.)	(Million EUR)	(Thousand tonnes)	Tonne
AT	4 689 420	104 663,3	26 487,8	405 241,4	55 477,8	13 943
BE	5 259 744	121 509,3	33 173,5	507 929,6	95 240,5	18 163
BG	3 268 751	11 616,6	10 164,4	71 060,1	32 757,1	30 572
HR	1 724 587	12 303,7	6 887,6	58 455,1	11 486,9	27 030
CY	666 826	4 850,2	1 585,0	24 927,6	6 789,9	991
CZ	5 274 427	61 964,2	25 315,2	238 249,5	105 661,3	24 391
DK	3 070 703	75 837,1	13 815,1	342 961,7	31 942,8	11 976
EE	702 973	9 090,4	2 789,7	31 169,0	12 955,1	4 840
FI	2 774 226	59 196,0	24 805,3	250 664,0	35 855,1	14 271
FR	31 271 173	612 198,0	138 965,3	2 502 118,0	297 194,0	189 218
DE	43 386 527	770 497,0	197 569,3	3 617 450,0	675 066,2	83 388
EL	4 574 730	24 164,9	14 911,9	181 500,4	51 902,7	35 524
HU	4 920 977	41 941,5	18 793,6	153 963,3	41 323,2	37 802
IE	2 551 560	96 941,3	11 061,6	434 069,7	43 826,3	12 675
IT	25 087 249	373 419,8	114 384,2	1 822 344,5	308 306,1	149 106
LV	956 869	7 461,4	3 980,3	33 348,9	8 209,6	17 760
LT	1 479 760	12 259,4	5 662,4	56 478,1	7 631,9	7 194
LU	333 257	13 156,8	3 440,4	72 360,9	7 815,9	1 214
MT	281 266	3 107,5	525,0	15 327,3	1 606,8	380
NL	9 653 277	184 405,0	43 215,0	870 587,0	144 394,9	14 196
PL	18 519 217	96 897,0	74 185,6	576 382,6	309 731,8	297 282
PT	5 190 888	43 639,5	15 772,0	216 053,2	33 228,1	45 490
RO	8 200 518	58 596,2	25 279,8	241 611,3	27 860,6	116 136
SK	2 777 251	19 254,6	10 508,3	100 255,7	27 500,8	18 609
SI	1 045 561	10 581,6	4 767,6	52 278,8	9 917,6	10 087
ES	23 384 158	245 709,0	78 607,5	1 222 290,0	185 287,5	135 005
SE	5 514 678	138 383,8	32 156,1	540 734,0	-4 765,1	15 907

efficiencies, none is dominated by a central point.

5. Conclusion. We have analysed the properties of path-based models with the AIP direction that was originally proposed to guarantee the feasibility and hence also the well-definedness of the super-efficiency DEA models. We have shown that even the DEA models for standard efficiency evaluation, defined using the AIP direction, have good properties. Thus, they can be used when measuring changes in productivity over time using the Malmquist or Luenberger indicator, where it is important to apply the same model to assess decision-making units from the inside as well as from the outside of the technology set. The advantage is that path-based models with the AIP direction can also be applied to negative data, which we demonstrated on the example of the eco-efficiency analysis of 27 EU countries.

Acknowledgments. The authors thank the anonymous referee for valuable comments.

REFERENCES

Table 4.2

Efficiencies, super-efficiencies and ranking for the standard and environmental DDF-g and HDF-g models with the AIP directions applied to 27 EU countries. The scores greater than one refer to the super-efficiency scores of strongly efficient units.

	Standard models				Environmental models			
	DDF		HDF		DDF		HDF	
DMU	Score	Rank	Score	Rank	Score	Rank	Score	Rank
AT	0.975	23	0.975	23	0.992	20	0.992	20
BE	0.984	20	0.984	20	0.985	24	0.985	24
BG	1.001	8	1.001	8	1.001	12	1.001	12
HR	0.997	14	0.997	14	0.998	18	0.998	18
CY	1	9	1	9	1	13	1	13
CZ	0.901	27	0.901	27	0.959	26	0.959	26
DK	0.997	16	0.997	16	1.003	11	1.003	11
EE	0.991	18	0.991	18	0.992	21	0.992	21
FI	0.99	19	0.99	19	0.991	22	0.991	22
FR	0.967	25	0.968	25	1.15	2	1.162	2
DE	1.31	1	1.449	1	1.31	1	1.449	1
EL	0.945	26	0.945	26	0.947	27	0.947	27
HU	0.98	22	0.98	22	0.983	25	0.983	25
IE	1.037	2	1.037	2	1.037	4	1.037	4
IT	1.008	4	1.008	4	1.01	7	1.01	8
LV	0.998	12	0.998	12	0.998	15	0.998	15
LT	0.997	15	0.997	15	0.998	19	0.998	19
LU	1.006	6	1.006	6	1.006	10	1.006	10
MT	1.007	5	1.007	5	1.008	9	1.008	9
NL	1	10	1	10	1.012	6	1.012	6
PL	1.022	3	1.022	3	1.022	5	1.022	5
PT	0.997	17	0.997	17	0.998	17	0.998	17
RO	0.981	21	0.981	21	0.991	23	0.991	23
SK	0.997	13	0.997	13	0.998	16	0.998	16
SI	0.998	11	0.998	11	0.999	14	0.999	14
ES	1.003	7	1.003	7	1.01	8	1.011	7
SE	0.973	24	0.973	24	1.082	3	1.082	3

[1] J. Aparicio, J. T. Pastor, and S. C. Ray, *An overall measure of technical inefficiency at the firm and at the industry level: The 'lost profit on outlay'*, European Journal of Operational Research, 226 (2013), pp. 154–162.

[2] 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.

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

[4] R. G. Chambers, R. Fāure, and S. Grosskopf, *Productivity growth in APEC countries*, Pacific Economic Review, 1 (1996), pp. 181–190.

[5] 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.

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

[7] 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.

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

[9] M. Halická and M. Trnovská, *Negative features of hyperbolic and directional distance mod-*

els for technologies with undesirable outputs, Central European Journal of Operations Research (CEJOR), 26 (2018), pp. 887–907.

[10] M. Halická, M. Trnovská, and A. Černỳ, *On indication, strict monotonicity, and efficiency of projections in a general class of path-based data envelopment models*, arXiv preprint arXiv:2311.16382, (2023).

[11] 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.

[12] K. Kerstens and I. Van de Woestyne, *Negative data in dea: a simple proportional distance function approach*, Journal of the Operational Research Society, 62 (2011), pp. 1413–1419.

[13] M. C. A. S. Portela, E. Thanassoulis, and G. Simpson, *Negative data in DEA: A directional distance approach applied to bank branches*, Journal of the Operational Research Society, 55 (2004), pp. 1111–1121.

[14] M. Trnovská and M. Halická, *Applying path-based models to negative data with the focus on super-efficiency*, (2023). https://www.researchgate.net/publication/375895714_Applying_path-based_models_to_negative_data_with_the_focus_on_super-efficiency.

[15] 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.

[16] 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.