

Revisiting the decomposition of cost efficiency for non-homothetic technologies: a directional distance function approach

Juan Aparicio¹ · José L. Zofío²

Published online: 18 August 2017
© Springer Science+Business Media, LLC 2017

Abstract In the early 1980's Kopp and Diewert proposed a popular method to decompose cost efficiency into allocative and technical efficiency for parametric functional forms based on the radial approach initiated by Farrell. We show that, relying on recently proposed homogeneity and duality results, their approach is unnecessary for self-dual homothetic production functions, while it is inconsistent in the non-homothetic case. By stressing that for homothetic technologies the radial distance function can be correctly interpreted as a technical efficiency measure, since allocative efficiency is independent of the output level and radial input reductions leave it unchanged, we contend that for non-homothetic technologies this is not the case because optimal input demands depend on the output targeted by the firm, as does the inequality between marginal rates of substitution and market prices—allocative inefficiency. We demonstrate that a correct definition of technical efficiency corresponds to the directional distance function because its flexibility ensures that allocative efficiency is kept unchanged through movements in the input production possibility set when solving technical inefficiency, and therefore the associated cost reductions can be solely—and rightly—ascribed to technical-engineering-improvements. The new methodology allowing for a consistent decomposition of cost inefficiency is illustrated resorting to simple examples of non-homothetic production functions.

Keywords Non-homotheticity · Technical efficiency · Allocative efficiency · Directional distance function

JEL Classification C61 · D21 · D24

1 Introduction

In 1982 Kopp and Diewert proposed the first methodology based solely on duality theory to decompose cost (in)efficiency into technical and allocative components for the case of parametric specifications. Their approach did not require resorting to the primal approach by direct or indirect estimation of a production function and its associated minimum cost share equations as previous proposals by Schmidt and Lovell (1979) or Kopp (1981), but simply knowledge of the cost function. A system of equations involving optimal demands for inputs—applying Shephard's Lemma—and relative input quantities allowed determination of the unknown reference technical efficient benchmark for any firm, as well as its associated input—shadow—price vector. Based on this solution a straightforward decomposition of cost efficiency into allocative and technical efficiency was possible. Kumbhakar and Lovell (2000) discuss Kopp and Diewert's (1982) original proposal, along with subsequent refinements by Zieschang (1983) and Mensah (1994), who improved the methodology by simplifying the system of equations to be solved, resulting in less computational requirements and numerical difficulties.

A common feature of all these and subsequent contributions, including those extending the previous analytical frameworks based on numerical solutions to an econometric setting that allows estimation of the cost function—

✉ Juan Aparicio
j.aparicio@umh.es

¹ Center of Operations Research (CIO), Universidad Miguel Hernández de Elche, 03202 Elche, Alicante, Spain

² Departamento de Análisis Económico: Teoría Económica e Historia Económica, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Kumbhakar (1997), is that they relied on Farrell's (1957) radial definition of technical efficiency, related both to the coefficient of resource utilization by Debreu (1951) and the inverse of Shephard's (1953) input distance function. Using Farrell's approach implies that technical efficiency is measured against the reference isoquant corresponding to the observed output level, regardless whether that output level was the one originally intended by the firm. Once the firm is projected onto that isoquant, and thanks to duality, allocative efficiency, defined as the difference between minimum cost and the cost at the technically efficient projection, is measured as a residual. As we show in this study, and despite the fact that the previous contributions allowed for both homothetic and non-homothetic technologies, Farrell's approach is only consistent for the former case, where radial reductions of inputs can be rightly interpreted as technical—engineering—improvements resulting in cost savings, because whatever the allocative efficiency magnitude resulting from the (in)equality of marginal rates of substitution with input price ratios, it does not change along the radial contraction path represented by the input distance function. This result stems from one remarkable technological property normally taken for granted in the literature by customarily assuming homotheticity, that the marginal rates of substitution among inputs are independent of the output level, and therefore the radial contractions of input quantities leave allocative efficiency unchanged.

This would not be relevant if one is willing to accept that the observed output level of the firm is the one targeted by its managers, and therefore optimal input demands are determined at that output level. However, adopting an input orientation to measure and decompose cost efficiency does not safeguard from the fact that the researcher does not know what is the output level originally intended by the firm. In the case of non-homothetic technologies, where the optimal input demands that minimize production cost depend on the output levels, allocative efficiency changes with the reference output level, and its residual nature is no longer consistent with a radial definition of technical efficiency. Consequently, the definition of allocative efficiency and its associated distance function counterpart—measuring technical efficiency—has gained recent attention. Bogetoft et al. (2006) were the first ones to propose a non-residual definition of allocative efficiency separate from its technical efficiency counterpart. They showed that under input homotheticity, allocative efficiency is independent of output level and that the order in which cost efficiency is decomposed, first technical and then allocative, or vice versa (i.e., the *reverse* approach), is irrelevant as they yield the same results. However, they failed to fully acknowledge that for non-homothetic technologies the radial input distance function cannot be truly considered as a correct measure of technical efficiency because it does not leave allocative

efficiency unchanged, as it depends on the output level that is considered to be measured.

This has been also recently stressed by Aparicio et al. (2015), who in a non-parametric non-homothetic Data Envelopment Analysis framework, show that a consistent definition of technical efficiency can only be achieved by resorting to the directional distance function¹—Chambers et al. (1996), whose flexibility allows to measure technical efficiency without altering allocative efficiency simultaneously. In short, they unveil that behind the apparently casual residual nature of allocative efficiency, a rationale for cost efficiency decomposition exists, and that indeed that rationale must be extended to non-homothetic technologies proposing a new model. Given the advantages of assuming homotheticity as the most common functional restriction—see Chambers and Mitchell (2001), it comes as no surprise that it is routinely imposed by researchers without testing for non-homotheticity, even if the latter case is the most common situation.²

In this study we revisit the decomposition of overall cost efficiency into technical and allocative components in the light of the standard and reverse approaches and extend the associated analytical framework to non-homothetic technologies. In this situation we show that the choice of reference output level is of paramount importance as both approaches are not equivalent, and introduce a new methodology that based on the definition of the directional distance function, allows us to reconcile both approaches and regain theoretical consistency under a general technological specification.³ In short, this is accomplished by ensuring that when measuring the technical efficiency of a firm producing in the interior of the production possibility set through movements to the frontier, allocative efficiency does not change along the process. This constitutes the desired condition that is unintentionally kept in the standard Farrell approach for homothetic technologies. Knowing that technologies will not generally exhibit the stylized homotheticity assumption, it is mandatory that when defining, interpreting and correctly decomposing cost efficiency, a correct measure of technical efficiency must keep constant its allocative efficiency counterpart. We introduce the necessary theoretical results to support our new

¹ See also Luenberger (1992) and his general notion of shortage function.

² From a parametric perspective, early confirmation of non-homothetic technologies is given by Christensen and Greene (1976) and Sato (1977), who show that for relevant electric utilities and manufacturing industries, non-homothetic specifications are statistically significant. In a non-parametric context, Aparicio et al. (2015) show that the most common DEA technologies are non-homothetic, except for the single output and constant returns to scale model.

³ As opposed to Kopp (1981, p. 490) who state that: “we can ignore homogeneity or homotheticity assumptions since our concern is on a single isoquant”.

methodology and illustrate how to solve the economic models and calculate and decompose cost efficiency for non-homothetic technologies. We do so analytically for well-defined production functions within the determinist framework initiated by Aigner and Chu (1968), who relied in mathematical programming, and that was extended to cost efficiency decomposition by the above mentioned authors. In this way we give continuity and revitalize a strand of literature that has not been revisited in the light of recent theoretical developments.⁴

The paper is organized as follows. In Section 2 we recall the standard and reverse Farrell approaches introduced by Bogetoft et al. (2006), showing that the radial distance function correctly characterizes technical efficiency as radial projections of inefficient firms leave allocative (in) efficiency unchanged, and therefore all cost savings can be attributed to technical efficiency gains. We illustrate our discussion with the standard Cobb-Douglas homothetic technologies, corresponding to an example by Kopp and Diewert (1982). Section 3 is devoted to introduce the correct decomposition of economic efficiency when the technology is non-homothetic by relying on the directional distance function as a measure of technical efficiency, and extends the previous rationale of ensuring that allocative (in)efficiency is kept constant as production cost is reduced through changes in the employed inputs quantities. We introduce the necessary programs to calculate the directional distance function that ensures a correct decomposition of cost inefficiency, and illustrate the model making use of non-homothetic Cobb-Douglas. Section 4 concludes.

2 Technical, allocative and costs efficiency with homothetic technologies: the standard approach

2.1 The production technology and technical efficiency

Let us consider the production possibility set $T = \{(x, y) : x \in R_+^N, y \in R_+^M, x \text{ can produce } y\}$. We assume that T satisfies the customary axioms, including closedness, Färe (1988). In this case, and for the single output case: $M = 1$, the technology can be represented by a production function $f: R_+^N \rightarrow R_+$, that is defined by:

$$f(x) = \max_y \{y : (x, y) \in T\}, \quad (1)$$

This definition stresses the envelope nature of the production function by characterizing the maximum amount of

output that can be obtained from any combination of inputs. The advantage of this interpretation is that it leaves room for technical inefficiency, since under the appropriate assumptions we can define the technology set departing from the production function as $T = \{(x, y) : f(x) \geq y, x \in R_+^N, y \in R_+\}$. We assume that: (i) the production function is well-behaved satisfying all desirable neoclassical properties and regularity conditions, particularly quasi-concavity, which ensures that the associated input production possibility sets are convex, Madden (1986); and (ii) it is continuous and twice differentiable.⁵ A relevant property is that the technology is homothetic when the production function is homogenous of degree $r > 0$: $f(\delta x) = \delta^r f(x)$. For $r > 1$, $r = 1$, and $r < 1$, the technology exhibits increasing, constant or decreasing returns to scale, respectively.⁶

Since we focus our analysis on the decomposition of economic efficiency into technical and allocative efficiency, it is better to represent the technology by means of the associated distance function, which provides a straightforward measure of the former. Assuming a cost minimizing behavior on the part of the firm, it is also convenient to define the input requirement set as $L(y) = \{x : (x, y) \in T\} = \{x \in R_+^N : f(x) \geq y\}$ —and therefore $T = \{(x, y) : x \in L(y), y \in R_+\}$. We say that the technology is homogeneous of degree r if $L(\delta y) = \delta^{1/r} L(y)$ (see Färe and Mitchell 1993 and Boussemart et al. 2009). Additionally, the input requirement set allows us to recall Shephard's input distance function as follows:

$$D_I(x, y) = \sup_{\lambda > 0} \{\lambda : (x/\lambda) \in L(y)\}. \quad (2)$$

This function completely characterizes the technology under weak disposability of inputs (Färe and Primont 1995), and allows us to determine the technical efficiency of any firm (x^1, y^1) as follows: Considering the observed output level y^1 as reference, we can define its associated isoquant as $\text{Isoq } L(y^1) = \{x : x \in L(y^1), \delta x \notin L(y^1) \forall \delta < 1\}$, $y^1 > 0$, and it can be shown that $x^1 \in \text{Isoq } L(y^1)$ if and only if $D_I(x^1, y^1) = 1$. Therefore, for $D_I(x^1, y^1) > 1$ the firm is technically inefficient, and its projection on the isoquant is denoted by $\hat{x}^1 = x^1 / D_I(x^1, y^1) \in \text{Isoq } L(y^1)$.

One may also think of inefficiency from the output perspective, and rely on the output distance function to

⁴ There are recent contributions that use semi-parametric and non-parametric techniques to estimate stochastic frontier models allowing for flexible functional forms—e.g. Delis et al. (2014), whose analytical framework could be used to implement the ideas introduced in this study.

⁵ The quasi-concavity assumption, ensuring that the input isoquants are convex, is satisfied by the most common functional forms—including those presented in subsequent sections; e.g., Cobb-Douglas. The regularity and differentiability conditions of the production function (1) passes on the distance functions defined below—see Blackorby and Diewert (1979).

⁶ Boussemart et al. (2009) introduced a more general definition in the literature for multi-output multi-input contexts: a production technology T is said to be homogeneous of degree α if for all $\lambda > 0$ $(x, y) \in T \Rightarrow (\lambda x, \lambda^\alpha y) \in T$. In particular, if $f(\delta x) = \delta^r f(x)$ for all $\delta > 0$ and $T = \{(x, y) : f(x) \geq y, x \in R_+^N, y \in R_+\}$, then T is homogeneous of degree r following Boussemart et al.'s definition.

equivalently determine if (x^1, y^1) is technically efficient or not. Define the output distance function as:

$$D_O(x, y) = \inf_{\theta > 0} \{ \theta : f(x) \geq y/\theta \}, \quad (3)$$

then if (x^1, y^1) is interior to T —technically inefficient, $D_O(x^1, y^1) < 1$, while it is efficient as long as $D_O(x^1, y^1) = 1$ and the inequality in (3) is an equality. In the event of (x^1, y^1) being inefficient, its projection in the output dimension corresponds to the following vector $(x^1, y^2) = (x^1, y^1/D_O(x, y))$, $y^2 > y^1$, showing that the observed input vector yields a higher amount of output. In this case it can be proved that the efficient projection belongs to the frontier of the technology: $y^2 = f(x^1)$, or alternatively, the isoquant defined as $\text{Isoq } L(y^2) = \{x : x \in L(y^2), \phi x \notin L(y^2) \forall \phi < 1\}$, with $x^1 \in \text{Isoq } L(y^2)$. In case that (x^1, y^1) were efficient, $D_O(x^1, y^1) = D_I(x^1, y^1) = 1$, and therefore $(x^1, y^1) = (\hat{x}^1, y^1) = (x^1, y^2)$. However, when the firm is technically inefficient it is verified that: $(x^1, y^1) \neq (\hat{x}^1, y^1) \neq (x^1, y^2)$.

2.2 The cost function and allocative efficiency

As previously discussed in the introduction, the implications of the previous result regarding the two alternative orientations that researchers use when measuring technical efficiency are quite relevant for the decomposition of cost efficiency so as to learn from its sources, particularly the correct definition of technical efficiency in terms of its allocative counterpart. As a cost minimizing firm demands input quantities taking as reference its *ex-ante* intended output level, but the researcher either observes *ex-post* actual production y^1 or, based on the excess of inputs consumed by the firm to produce it, can determine the potential output level y^2 that would be attainable, the choice of orientation is quite relevant when assessing allocative efficiency. Therefore, when optimal input quantities at y^1 or y^2 differ from those actually observed x^1 , or those at the technically efficient projection \hat{x}^1 , the firms incur in allocative inefficiency, whose magnitude may vary at both output levels. In this section, we recall Bogetoft et al.'s (2006) results showing that for the case of homothetic technologies they are the same, and therefore it does not constitute a concern for the analyst, contrarily to the non-homothetic case that we address in the next section.

We elaborate this issue formally by defining the cost function—frontier—dual to the technology as $C(y, w) = \min_x \{wx | x \in L(y)\}$ ⁷, representing the minimum cost of

⁷ Notice that $L(y)$ is closed since we are supposing that T is closed. However, it is not enough to assure that 'inf' can be substituted by 'min' in $\inf_x \{wx | x \in L(y)\}$. So, hereafter, we assume that the optimization problem associated with the calculation of the cost function $C(y, w)$ always attains its minimum in the set $L(y)$. There exist several sufficient conditions in the literature which assure such result. For example, Shephard (1970, p. 223) assumed that the subset of Pareto-efficient points of $L(y)$ is bounded. Another case is when the technology is a polyhedral set (see Mangasarian 1994, p. 130).

producing y given the vector of inputs prices $w = (w_1, \dots, w_N)$, and satisfying the usual properties, including continuity and second order differentiability. Among these, we recall here that when the single output technology represented by the production function is homogenous of degree r , the cost function is separable and can be expressed as follows: $C(y, w) = y^{1/r} C(1, w)$ —Silberberg and Suen (2000, p. 229).

Given a particular vector of market prices, w , we can recover the amount of inputs minimizing the cost of production by way of Shephard's Lemma, i.e.,

$$x^*(y, w) = \nabla_w C(y, w), \quad (4)$$

where $\nabla_w C(y, w) \equiv [\partial C(y, w)/\partial w_1, \dots, \partial C(y, w)/\partial w_N]$. This vector of input demand equations solves the first order conditions that result from the cost minimizing problem subject to the production function constraint, which is contingent on a particular output level y , and implying that $x^*(y, w) \in \text{Isoq } L(y)$. Again, if the production function is homogenous of degree r , the system of demand equations can be expressed as:

$$x^*(y, w) = y^{1/r} \nabla_w C(1, w). \quad (5)$$

For any two inputs k and l with associated market prices w_k and w_l , the first order conditions also imply that the marginal rate of technical substitution of factor k for factor l must be equal to the price ratios⁸:

$$MRS_l^k = -dl/dk = f_k(x)/f_l(x) = w_k/w_l, \quad (6)$$

where $f_k(x) = \partial f(x)/\partial x_k$ and $f_l(x) = \partial f(x)/\partial x_l$ are marginal productivities. With this information we define the (input-oriented) allocative efficiency of (x^1, y^1) in the following way:

$$AE(x^1, y^1, w) = \frac{C(y^1, w)}{wx^1} = \frac{C(y^1, w)}{wx^1/D_I(x^1, y^1)} = \frac{w \nabla_w C(y^1, w)}{wx^1/D_I(x^1, y^1)} = \frac{wx^*(y^1, w)}{wx^1/D_I(x^1, y^1)}, \quad (7)$$

which under homogeneity of degree r can be expressed as follows:

$$AE(x^1, y^1, w) = \frac{(y^1)^{1/r} C(1, w)}{wx^1/D_I(x^1, y^1)} = \frac{(y^1)^{1/r} w \nabla_w C(1, w)}{wx^1/D_I(x^1, y^1)} = \frac{(y^1)^{1/r} wx^*(1, w)}{wx^1/D_I(x^1, y^1)}, \quad (8)$$

As for the values of $AE(x^1, y^1, w)$, if $\hat{x}^1 = x^1/D_I(x^1, y^1) = x^*(y^1, w)$ the firm is allocative efficient with $AE(y^1, x^1, w) = 1$, and the marginal rates of substitution are equal

⁸ We assume that given our assumptions about the production technology, the second order conditions are verified and therefore the sign of the bordered Hessian determinant is negative.

to the input price ratios, Eq. (6). It follows immediately that if $x^1/D_I(x^1, y^1) \neq x^*(y^1, w)$ the firm is allocative inefficient with $AE(x^1, y^1, w) < 1$, with the marginal rates of substitution differing from relative prices.

Following now the *reverse* approach proposed by Bogetoft et al. (2006), we define the allocative efficiency associated to the optimal input demands if the firm were to take as reference output level y^2 with $x^1 \in \text{Isoq } L(y^2)$ —and therefore $D_I(x^1, y^2) = 1$ so the firm is technically efficient and the second equality below holds. In that case:

$$AE(x^1, y^2, w) = \frac{C(y^2, w)}{wx^1/D_I(x^1, y^2)} = \frac{w\nabla_w C(y^2, w)}{wx^1} = \frac{wx^*(y^2, w)}{wx^1}, \quad (9)$$

which under degree r homogeneity can be expressed equivalently as:

$$\begin{aligned} AE(x^1, y^2, w) &= \frac{(y^2)^{1/r} C(1, w)}{wx^1/D_I(x^1, y^2)} = \frac{(y^2)^{1/r} w\nabla_w C(1, w)}{wx^1} \\ &= \frac{(y^2)^{1/r} wx^*(1, w)}{wx^1}. \end{aligned} \quad (10)$$

This expression is numerically interpreted in the same way as above.

We now recall that homogeneity allows the following definition of the input distance function.

Lemma 1 [Färe and Mitchell 1993]. Let f be a production function homogeneous of degree $r > 0$, then $D_I(x, y) = y^{-1/r} D_I(x, 1)$.

Proof This result is a direct consequence of Proposition 1 (iii) in Färe and Mitchell (1993) when $M = 1$. ■

Relying on Boussemart et al. (2009), we recall that, for homogenous technologies, the following relationship between the input and output distance functions holds:

Proposition 1 [Boussemart et al. 2009]. Let f be a production function homogeneous of degree $r > 0$, then $D_I(x, y) = \frac{1}{D_O(x, y)^{1/r}}$ and $D_O(x, y) = \frac{1}{D_I(x, y)^r}$.

Proof It follows directly from Proposition 3.3(a) in Boussemart et al. (2009). ■

We can now establish the following result relating the values of the allocative efficiency measures $AE(x^1, y^1, w)$ and $AE(x^1, y^2, w)$:

Proposition 2 Let f be a production function homogeneous of degree $r > 0$, then (i) the allocative efficiency of a firm is the same regardless of the output level that is taken as benchmark for its input demands: $AE(x^1, y^1, w) = AE(x^1, y^2, w)$, and (ii) the marginal rates of substitution are independent of the output level and remain constant along

the ray vector associated to the input distance function: $MRS_I^k = -dl/dk = f_k(x^1)/f_l(x^1) = f_k(x^1/\tau)/f_l(x^1/\tau)$ for all $1 \leq \tau \leq D_I(x^1, y^1)$.

Proof (i) If the technology is homogeneous of degree r then (i) $AE(x^1, y^1, w^*) = (y^1)^{1/r} C(1, w)/wx^1/D_I(x^1, y^1)$ [see (8)] and $AE(x^1, y^2, w) = (y^1/D_O(x^1, y^1))^{1/r} C(1, w)/wx^1$ by (10) with $y^2 = y^1/D_O(x^1, y^1)$. Therefore $AE(x^1, y^2, w)/AE(x^1, y^1, w) = [((y^1)^{1/r}/D_O(x^1, y^1)^{1/r})(C(1, w)/wx^1)]/[(y^1)^{1/r} C(1, w)/(wx^1/D_I(x^1, y^1))] = (1/D_O(x^1, y^1)^{1/r})/D_I(x^1, y^1) = 1$, since $1/D_O(x^1, y^1)^{1/r} = D_I(x^1, y^1)$ by Proposition 1; (ii) Making use of the property that the first order partial derivatives of a homogenous production function of degree r in the input quantities are homogenous of degree $r - 1$ (e.g., Silberberg and Suen 2000, p. 225), i.e., $\delta^{r-1} f_k(x) = f_k(\delta x)$, $\delta > 0$, we observe that $MRS_I^k = -dl/dk = f_k(x^1)/f_l(x^1) = f_k(x^1/\tau)/f_l(x^1/\tau)$. ■

As a counterpart to the cost function separability under linear homogeneity of degree r : $C(y, w) = y^{1/r} C(1, w)$, and for future reference, we recall its associated input homotheticity counterpart in terms of the input requirement set: $L(y) = y^{1/r} L(1)$ (see Färe and Mitchell 1993). Additionally, under the hypothesis of homogeneity it is possible to prove that two firms located on the same output isoquant for y^2 share the same input-oriented technical efficiency when it is calculated with respect to a common level y^1 .

Lemma 2 Let f be a production function homogeneous of degree $r > 0$. Let $x^1, x^2 \in \text{Isoq}(y^2)$. Then, $D_I(x^1, y^1) = D_I(x^2, y^1)$.

Proof By definition, $\hat{x}^1 = \frac{x^1}{D_I(x^1, y^1)}$. In this way, $x^1 = D_I(x^1, y^1) \hat{x}^1$, which implies that $f(x^1) = f(D_I(x^1, y^1) \hat{x}^1) = D_I(x^1, y^1)^r f(\hat{x}^1)$ using that $f(x)$ is homogeneous. Now, applying that $f(x^1) = y^2$ and $f(\hat{x}^1) = y^1$, we have that $D_I(x^1, y^1) = \left(\frac{y^1}{y^2}\right)^{1/r}$. Finally, since $x^1, x^2 \in \text{Isoq}(y^2)$, we obtain that $D_I(x^1, y^1) = D_I(x^2, y^1)$. ■

As we recall later, it follows from this result that, under the homotheticity assumption, the values of Shephard's input distance function for any two firms belonging to a given isoquant $\text{Isoq } L(y^2)$ are equal when they are projected onto the same reference isoquant $\text{Isoq } L(y^1)$ —Bogetoft et al. (2006). From a graphical perspective the homotheticity relationship $L(y) = y^{1/r} L(1)$, implies that the inputs sets are “parallel” along a given ray vector.

2.3 Decomposing cost efficiency into allocative and technical terms

With this background, we define the cost efficiency of firm (x^1, y^1) and decompose it into its allocative and technical

components. The overall cost efficiency when firms demand inputs taking as reference Isoq $L(y^1)$ is given by:

$$\begin{aligned} CE(x^1, y^1, w) &= \frac{C(y^1, w)}{wx^1} = \frac{C(y^1, w)}{w(x_k^1/D_I(x^1, y^1))} \frac{1}{D_I(x^1, y^1)} \\ &= AE(x^1, y^1, w) TE(x^1, y^1) \end{aligned} \quad (11)$$

corresponding to the standard approach proposed by Farrell. Following Bogetoft et al. (2006) we can also define the *reversed* overall cost efficiency taking Isoq $L(y)$ as the reference for the allocative efficiency evaluation—again assuming $x^1 \in \text{Isoq } L(y^2)$ so the firm is technically efficient:

$$\begin{aligned} CE^R(x^1, y^1, y^2, w) &= AE^R(x^1, y^2, w) \cdot TE^R(y^1, y^2, w), \\ AE^R(x^1, y^2, w) &= AE(x^1, y^2, w) = \frac{C(y^2, w)}{wx^1}, \end{aligned} \quad (12)$$

$$TE^R(y^1, y^2, w) = 1/D_I(x^*(y^2, w), y^1).$$

In their method, they first correct for allocative efficiency by changing the observed input bundle from x^1 to $x^*(y^2, w) = \arg \min \{wx : x \in L(y^2)\}$ on Isoq $L(y^2)$, and then remove technical inefficiency by reducing the input quantities radially from $x^*(y^2, w)$ to $x^*(y^2, w)\lambda$ on Isoq $L(y^1)$, where $\lambda = D(x^*(y^2, w), y^1)$. As a result of their Proposition 1, Bogetoft et al. (2006) show that under input homotheticity both decompositions are equivalent; i.e., $CE^R(x^1, y^1, y^2, w) = CE(x^1, y^1, w)$, $AE^R(x^1, y^2, w) = AE(x^1, y^1, w)$ and $TE^R(y^1, y^2, w) = TE(x^1, y^1)$. Here we summarize this result in terms of the input and output distance functions through the following result, since the equivalence of the allocative components was stated in Proposition 2.

Proposition 3 *Let f be a production function homogeneous of degree $r > 0$, then $TE^R(y^1, y^2, w) = TE(x^1, y^1) = 1/D_I(x^1, y^1) = D_O(x^1, y^1)^{1/r}$.*

Proof It follows directly from Proposition 1 in Bogetoft et al. (2006) and Proposition 1 above by Boussemart et al. (2009). ■

As a result of Propositions 2 and 3 we obtain the following:

Corollary 1 *Let f be a production function homogeneous of degree $r > 0$, the input distance function can be considered as a measure of technical efficiency because the radial contraction that it brings along the ray vector (factor beams along which factors are held in fixed proportions) leaves the marginal rates of substitution unchanged, as they do not depend on the output levels, and therefore the*

difference between the marginal rates of substitution and the price ratios remains constant: $MRS_I^k(x_k^1, x_l^1) = MRS_I^k(x_k^1/D_I(x^1, y^1), x_l^1/D_I(x^1, y^1)) \geq w_k/w_l$.

This implies that regardless of whether the firm is allocative efficient ($MRS_I^k(\cdot) = w_k/w_l$) or not ($MRS_I^k(\cdot) \neq w_k/w_l$), the allocative efficiency level given by (7), which is independent of the output level or input isoquant that is chosen as reference, is constant along the ray vector, and therefore any cost inefficiency reduction through radial projections can be attributed solely to technical inefficiency reductions. That is, the input distance function can be consistently used to determine the technical efficiency level regardless of the intended output level that the managers of a firm might have targeted, and therefore taken as benchmark when planning their input demands.

2.4 Homothetic Cobb-Douglas production function: H-CD

For the purpose of illustration we revisit this classic three-input H-CD production function: $y = 0.049x_1^{0.25}x_2^{0.1}x_3^{0.7}$, which is homogeneous of degree $r = 1.05$ and whose dual cost frontier is $C(y, w) = y^{1/r}C(1, w) = y^{1/1.05}40.4w_1^{0.238}w_2^{0.095}w_3^{0.667}$ Kopp and Diewert's (1982). Given the firm under evaluation $(x^1, y^1) = (x_1^1, x_2^1, x_3^1, y^1) = (500, 50, 16,2754, 1300)$, input prices: $w^1 = \$12,000$, $w^2 = \$20,000$, $w^3 = \$65$, and optimal input demands—applying Shephard's lemma: $(x_1^*(y, w), x_2^*(y, w), x_3^*(y, w)) = (290.6, 69.7, 15,0163.9)$, minimum cost is: $C(y^1, w) = \$14,641,987.2$ (see Table 1). Therefore cost efficiency is $CE(x^1, y^1, w) = 0.8325$, $AE(x^1, y^1, w) = 0.9689$, and $TE(x^1, y^1) = 1/D_I(x^1, y^1) = 0.8596$. Note that thanks to Proposition 1, since $D_O(x^1, y^1) = 1300/1523.73 = 0.8532$, we can recover $D_I(x^1, y^1) = 1/0.8532^{1/1.05} = 1.1633$, and therefore project the evaluated firm to the isoquant $y^1 = 1300$ without solving the system of equations proposed by Kopp and Diewert (1982), Zieschang (1983) or Mensah (1994), thereby rendering their method unnecessary for the case of homothetic self-dual technologies; i.e., the technically efficient input vector corresponds to $(\hat{x}^1, y^1) = (x^1/D_I(x^1, y^1), y^1) = (429.8, 43.0, 13,9910.8, 1300)$, with an associated cost of \$15,111,724.7.⁹ Consequently, since the radial projections associated to the restrictions keeping inputs' proportions (mix) constant are inconsistent in the case of non-homothetic technologies—as we show in what follows, their whole proposal to decompose cost efficiency is either unnecessary or incorrect, and should be discarded by

⁹ There are numerical discrepancies for the technically efficient input vector as a result of the iterative computational algorithm used by Kopp and Diewert (1982), who report the following values: (423,42, 13,1839). It is however easy to check that the output value corresponding to this vector given the production function is $y = 1278.4$, falling short from the reference value $y^1 = 1300$.

Table 1 Cost efficiency decomposition with a H -CD technology: $y = 0.049x_1^{0.25}x_2^{0.1}x_3^{0.7}$

	Input x_1	Input x_2	Input x_3	Cost efficiency (CE)	Allocative efficiency (AE)	Technical efficiency (TE)	\hat{x}_1	\hat{x}_2	\hat{x}_3
$L^H(y^2 = 1523.72)$ $C(y^2, w) = 17,032,589$	500.00	50.00	162,754.00	0.8329	0.9689	0.8596	429.82	42.98	139,910.73
	338.00	81.12	174,681.36	0.8596	1.0000	0.8596	290.56	69.73	150,164.03
	716.14	20.04	163,130.12	0.7471	0.8691	0.8596	615.63	17.23	140,234.14
$L^H(y^3 = 1750.00)$ $C(y^3, w) = 19,433,343$	570.48	57.05	185,694.32	0.7300	0.9689	0.7534	429.82	42.98	139,910.76
	385.64	92.55	199,302.80	0.7534	1.0000	0.7534	290.56	69.73	150,164.03
	817.08	22.86	186,123.51	0.6548	0.8691	0.7534	615.63	17.23	140,234.14
$L^H(y^4 = 2000.00)$ $C(y^4, w) = 22,068,761$	647.84	64.78	210,876.94	0.6428	0.9689	0.6635	429.82	42.98	139,910.77
	437.94	105.10	226,330.90	0.6635	1.0000	0.6635	290.56	69.73	150,164.03
	927.89	25.96	211,364.24	0.5766	0.8691	0.6635	615.63	17.23	140,234.14

Selected firms targeting different output levels larger than $L(y^1 = 1300)$. $C(y^1, w) = 14,641,987.26$, KD (1982)

researchers. As for the reverse Farrell approach, and exemplifying Proposition 2, it yields the same results for the three firms falling short from producing $y^2 = 1523.72$. The results for two additional sets of firms producing $y^3 = 1750$ and $y^4 = 2000$ are also presented in Table 1.

3 Technical, allocative and cost efficiency with non-homothetic technologies: the generalized approach

In this section we show that when the technology is non-homothetic, using the observed output level as reference benchmark to assess allocative efficiency as implied by the standard Farrell approach results in a wrong decomposition of cost efficiency if the firm had targeted an alternative output level. We show that the previous results summarized in Propositions 2 and 3 do not hold and therefore the standard and reverse Farrell approaches to evaluate the sources of cost inefficiency do not coincide, so a new general approach to decompose cost efficiency is introduced. First, we illustrate the inconsistency of the standard decomposition empirically by relaying on a non-homothetic version of the previous example and, secondly, we introduce the theoretical results on which the new generalized approach is based.

To ease the illustration we rely on a non-homothetic version of the standard Cobb-Douglas specification. Particularly, since we have to decide on a specific functional form, we rely on Sato (1977) and adopt the following specification:

$$f(x, y) = \prod_{k=1}^N x_k^{c_k(y)}, \quad (13)$$

which can be expressed more conveniently by taking logarithms and normalizing by $c_k(y)$ as $f(x, y) = \ln x_k +$

$\sum_{l=1}^{N-1} c_l(y) \ln x_l - h(y) = 0$, or, equivalently—see Sato (1975):

$$h(y) = \ln x_k + \sum_{l=1}^{N-1} c_l(y) \ln x_l. \quad (14)$$

Relevant for our analysis are the marginal rate of substitutions associated with (14), which for any two inputs k and l is:

$$MRS_l^k = -dl/dk = f_k(x^1)/f_l(x^1) = (x_l^1/x_k^1)g(y), \quad (15)$$

As opposed to the H -CD, the marginal rates of substitution for the NH -CD specification are variable at a constant factor ratio by depending on the output level, whose functional specification $g(y)$ is to be defined. Particularly, we assume the simplest specification for the output level function: $c_l(y) = \delta y$, where δ represents the non-homotheticity parameter. Assuming this specification we relate it to $h(y)$ in Eq. (14) so as to explicitly express y as a function of the inputs.¹⁰ To make the argument concrete we adopt the following formulation: $h(y) = 2^{y-1}$, $c(y) = 0.5y$, and consider the simplest non-homothetic extension of the two-inputs Cobb-Douglas specification found in any textbook, e.g., Silberberg and Suen (2000). With these functions the non-homothetic Cobb-Douglas NH -CD specification becomes $2^{y-1} = x_1^{0.5}x_2^{0.5y}$, while Eq. (14) is $(y-1)\ln 2 = 0.5\ln x_1 + 0.5y\ln x_2$. We can recover analytically the explicit expression for y :

$$y = \frac{\ln 2 + 0.5\ln x_1}{\ln 2 - 0.5\ln x_2}, \quad (16)$$

and the marginal rate of substitution corresponds to $MRS_2^1 = x_2^1/x_1^1y$. Therefore $g(y) = 1/y$, and as production

¹⁰ As $c_l(y)$ and $h(y)$ are arbitrary functions on a *a priori* basis, the non-homothetic expression of the production function is implicit generally.

varies along a given ray vector x_2^1/x_1^1 , so does the *MRS* at the rate specified by $\partial MRS_2^1/\partial y = \partial g(y)/\partial y = -1/y^2$. The—non-separable—dual cost frontier is

$$C(y, w) = w_1 \left(\frac{2^{y-1} w_2^{0.5y}}{w_1^{0.5y} y^{0.5y}} \right)^{1/0.5+0.5y} + w_2 \left(\frac{2^{y-1} w_1^{0.5y}}{w_2^{0.5y}} \right)^{1/0.5+0.5y},$$

where $(x_1^*(y, w), x_2^*(y, w)) = \left(\left(\frac{2^{y-1} w_2^{0.5y}}{w_1^{0.5y} y^{0.5y}} \right)^{1/0.5+0.5y}, \left(\frac{2^{y-1} w_1^{0.5y}}{w_2^{0.5y}} \right)^{1/0.5+0.5y} \right)$ are the input demand equations that

can be obtained, once again, applying Shephard's lemma. From the former conditions the expansion path for the *NH-CD* is $x_2^* = (w_1/w_2)x_1^*$.

In Fig. 1 we represent the map of isoquants for *NH-CD* along with the isocost function $C = w_1x_1 + w_2x_2$, and input prices $w_1 = w_2 = 1$. Considering initially $(x^1, y^1) = (x_1^1, x_2^1, y^1) = (4, 1, 1)$ as the evaluated firm, and the cost minimizing inputs' quantities for the observed output level $y^1 = 1$, $(x^*, y^1) = (x_1^*, x_2^*, y^1) = (1, 1, 1)$, we see that from (11) $CE(x^1, y^1, w^*) = 2/5 = 0.4$, $AE(x^1, y^1, w^*) = 2/2.5 = 0.8$, and $TE(x^1, y^1) = 1/D_I(x^1, y^1) = 1/2 = 0.5$, where $D_I(x^1, y^1) = x_1^{0.5} x_2^{0.5y} / 2^{y-1}$. Consequently, the technically efficient projection of the evaluated firm is: $(\hat{x}_1^1, \hat{x}_2^1, y^1) = (x_1^1/D_I(x^1, y^1), x_2^1/D_I(x^1, y^1), y^1) = (2, 0.5, 1)$, whose $MRS_2^1(x_1^1/D_I(x^1, y^1), x_2^1/D_I(x^1, y^1)) = x_2^1/D_I(x^1, y^1)/x_1^1/D_I(x^1, y^1) = 0.5/2 = 0.25$ differs from the price ratio $w_1/w_2 = 1$, and does not belong to the expansion path $x_2^* = (w_1/w_2)x_1^*$.

However, let us now consider the cost efficiency decomposition with respect to the isoquant $\text{Isoq } L^{NH}(y^2 = y^1/D_O(x^1, y^1) = 2)$ with $D_O(x^1, y^1) = 0.5$, taking as a reference Eq. (12). Assuming that the firm falls short of the intended output, $(x^1, y^2) = (x_1^1, x_2^1, y^2) = (2, 2, 2)$, its costs efficiency if we were to consider $\text{Isoq } L^{NH}(y^1 = 1)$ as reference would still be $CE(x^1, y^1, w) = 2/5 = 0.4$, but its allocative efficiency is now $AE(x^1, y^2, w) = 3/5 = 0.6$, while its technical efficiency based on Shephard's input distance function is $TE^R(y^1, y^2, w) = TE(x^1, y^1) = 1/D_I(x^1, y^1) = 1/2 = 0.5$. Consequently, opposed to the case of homothetic technologies, unknowing whether the firm's targets output or inputs so as to minimize production costs when the technology is non-homothetic results in an inconsistent decomposition of cost efficiency into the allocative and technical terms. Clearly, as $AE(x^1, y^2, w) = 0.6 < AE(x^1, y^1, w) = 0.8$, $CE(x^1, y^1, w) \neq AE^R(x^1, y^2, w) \cdot TE^R(y^1, y^2, w)$; i.e., Proposition 2 does not hold, and bridging the gap requires the introduction of an additional term that Bogetoft et al. (2006) named “second order” allocative efficiency, $AAE(x^1, y^2, w)$. This term emerges unnaturally as a consequence of relying on Shephard's radial distance function as definition of technical efficiency. As a result, for non-homothetic technologies, these authors propose the following decomposition of the reverse Farrell's approach taking $\text{Isoq } L^{NH}(y = 2)$ as reference for the allocative

efficiency evaluation.

$$\begin{aligned} CE(x^1, y^1, w) &= \frac{C(y^1, w)}{wx^1} \\ &= \frac{C(y^2, w)}{wx^1} \frac{1}{D_I(x^*(y^2, w), y^1)} \frac{wx^*(y^1, w)}{wx^*(y^2, w)/D_I(x^*(y^2, w), y^1)} \\ &= AE^R(x^1, y^2, w) TE^R(y^1, y^2, w) AAE(x^1, y^2, w). \end{aligned} \quad (17)$$

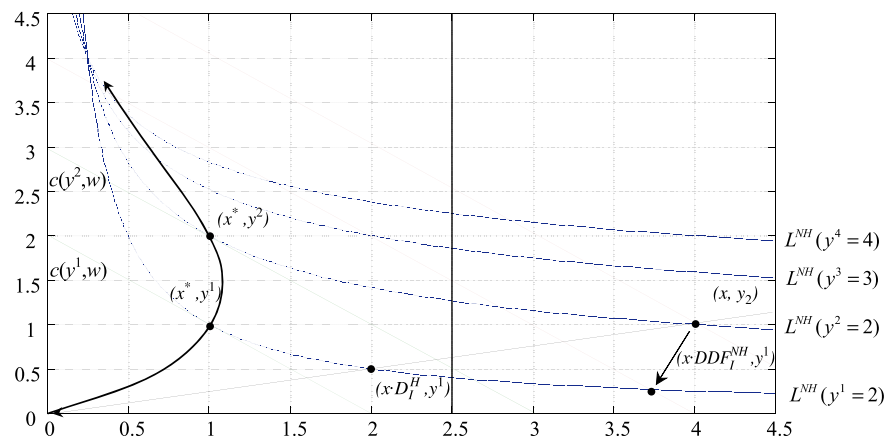
The definition of $AAE(x^1, y^2, w^*)$ as a residual closing the decomposition led Bogetoft et al. (2006) to conclude that homotheticity is required so as to get consistent measures of the allocative and technical efficiency terms; i.e., the criterion by which the standard and reverse Farrell's approaches coincide as summarized in Proposition 2 and Corollary 1. However, this did not prompted them to drop Shephard's input distance function and look for an alternative. This shows how embedded is the radial definition of technical efficiency after 50 years under Farrell's (1957) classic paradigm. The need for that second term residual, inconsistent with duality theory, exemplifies why a general approach based on the flexible directional distance function is required.¹¹

3.1 The directional distance function as efficiency measure for non-homothetic technologies

We rely on Chambers et al. (1996) to show that a complete and consistent decomposition of the standard and reverse approaches for non-homothetic technologies can be obtained without resorting to ad-hoc closing terms such as $AAE(x^1, y^2, w)$ in (17). As it is well-known, the directional input distance function generalizes Shephard's input radial distance function while preserving the desirable duality with the support cost function that enables a consistent decomposition into the usual technical and allocative components.

¹¹ There is yet a more drastic approach to this issue that considers the decomposition of technical and allocative efficiency essentially subjective. Zofio et al. (2013) discuss the flexibility of the directional distance function as a measure of technical efficiency in an exogenously pre-specified direction, and the arbitrary decomposition of cost efficiency into technical and allocative terms in which it results, as the former would change in value at the researcher's will, depending on the choice of direction. Consequently, being the allocative efficiency a residual, its value will also change accordingly. Their conclusion is that the use of flexible directional distance functions renders the cost efficiency decomposition exercise meaningless, as technical efficiency depends on the choice of the directional vector, unless a criterion is imposed; i.e., in the case of homothetic technologies radial contractions do keep allocative efficiency unchanged, so Shephard's distance function can be rightly considered as a measure of technical efficiency. For non-homothetic technologies, we use the same criterion of keeping allocative efficiency constant in the standard and reverse approaches, resulting in the same technical inefficiency values between isoquants.

Fig. 1 Cost efficiency decomposition with non-homothetic technologies: $h(y) = (x_1^{0.5} x_2^{(y)}); \quad 2^{y-1} = x_1^{0.5} x_2^{0.5y} (NH-CD)$



Let $g = \{g_1, \dots, g_N\}$ be a vector such that $g \neq 0_m$, then the directional input distance function defines as $\vec{D}_I(x, y; g) = \sup\{\beta : x - \beta g \in L(y)\}$. It can be proved that if $g = x^1$ then $\vec{D}_I(x^1, y^1; x^1) = 1 - 1/D_I(x^1, y^1)$ —e.g., Boussemart et al. (2010). We define cost inefficiency as $CI(x^1, y^1, w; g) = \frac{wx^1 - C(y^1, w)}{wg}$, and considering $\vec{D}_I(x^1, y^1; g)$ as the associated of technical inefficiency, the following Mahler inequality based on the dual relationship holds (see Färe and Grosskopf 1997 and Chambers et al. 1998):

$$\underbrace{\frac{wx^1 - C(y^1, w)}{wg}}_{\text{Cost Inefficiency (CI)}} \geq \underbrace{\vec{D}_I(x^1, y^1; g)}_{\text{Technical Inefficiency (TI)}}. \quad (18)$$

From the *standard* (Farrell) approach perspective, allocative inefficiency may be computed from (18) as a residual: $AI(x^1, y^1, w; g) = CI(x^1, y^1, w; g) - TI(x^1, y^1; g)$. Note that (i) the decomposition is additive instead of multiplicative given the nature of each directional and radial distance functions¹², and (ii) the change in denomination from efficiency to inefficiency as a result of the change in the efficient values: one for the radial distance function and zero for the directional distance function—the higher the value the better in the former, and the worse in the latter.

It can be remarked that although (18) offers flexibility in measuring cost and technical inefficiency through a wide array of reference vectors g , most practitioners resort to $g = x^1$ as their only choice. The reason is that when $g = x^1$ the overall additive cost inefficiency decomposition (18) is equivalent to the multiplicative setting based on the Shephard's input distance function and, therefore, the technical inefficiency values can be readily converted in equiproportional reductions in inputs, collapsing into the prevailing Farrell paradigm; i.e., the flexibility of g is not exploited. However, we have exemplified that this latter measure is inadequate to characterize technical efficiency in a non-

homothetic setting as the allocative (in)efficiency of an evaluated firm is not preserved through radial contractions. This limitation can be precisely overcome thanks to the flexibility of the directional distance function as long as we can consistently extend the notion of cost efficiency decomposition to non-homothetic technologies and, in doing so, provide a first theoretical extension to parametric models used empirically.

Once the definition of the standard overall economic efficiency decomposition in terms of the directional distance function has been presented, we extend Bogetoft et al.'s (2006) *reverse* approach to the case of the directional input distance function. As already presented, these authors introduced a way to estimate the “starting” allocative efficiency of (x^1, y^1) , which does not need the initial projection of this vector to the input requirement set defined by the observed output vector y^1 , Isoq $L(y^1)$. We contend that if one is interested in measuring the technical efficiency corresponding to x^1 in the input space by means of movements to the frontier associated with the production of y^1 , then one would ideally make sure that the allocative efficiency does not change along this process, as illustrated with the *NH-CD* case. Otherwise, one may not be sure that the cost savings derived from these input adjustments are consequence of exclusively technical—engineering—issues not related to allocative efficiency; i.e., changes in the inputs' proportions (mix) in the homothetic case.

The first step to obtain an input directional distance function requires the explicit definition of allocative efficiency with respect to Isoq $L(y^2)$ as the normalized difference between the minimum cost of producing y^2 given input prices w , and the observed cost for x^1 :

$$AI(x^1, y^2, w; g) = \frac{w(x^1 - \vec{D}_I(x^1, y^2; g)g) - C(y^2, w)}{wg} \quad (19)$$

$$= \frac{wx^1 - C(y^2, w)}{wg},$$

¹² See Aparicio et al. (2017).

where the second equality in (19) holds as long as $\bar{D}_I(x^1, y^2; g) = 0$ given that $x^1 \in \text{Isoq } L(y^2)$.

We can now retake the idea of correctly interpreting and measuring technical efficiency with non-homothetic technologies, so as to keep allocative efficiency constant along projections of the observed input vector x^1 . Then the following question arises: Is there a reference vector g so the input directional distance function actually measures technical inefficiency by leaving allocative efficiency unchanged? The answer is yes. Under the assumption of input homotheticity, the directional input distance function with reference vector $g = x^1$, always satisfies that the “starting” allocative inefficiency, measured with respect to $\text{Isoq } L(y^2)$; i.e., before projecting the original input vector x^1 to the isoquant of y^1 , and the “final” allocative inefficiency, after projecting the original input vector, are the same (see Aparicio et al. 2015, p. 886). Being allocative efficiency constant regardless the choice of decomposition, several remarks are in order: (1) The significance of Propositions 2 and 3 is that the *true* technical efficiency can be computed from the observed quantity data using the directional input distance with $g = x^1$, *even without knowing actual* input prices; (2) it turns out that the unrealized—but actually planned—output level y^2 that the inefficient firm might had targeted from an engineering perspective, does not need to be known by the researcher when determining the allocative efficiency, as its value is the same regardless that output level; (3) Moreover, as presented in Lemma 2 these distance functions have the same value for all firms belonging to the same output isoquant.

These results simplify the evaluation process of the firms in terms of the information required to correctly decompose cost inefficiency, as knowledge of the input prices, and the actually planned output level are unnecessary to correctly estimate technical efficiency (market prices are nevertheless needed to calculate cost efficiency).

3.2 Solving for the directional vector in the case of parametric non-homothetic technologies

Once the criterion of constant allocative efficiency for a correct measurement of technical efficiency with non-homothetic technologies has been established, which results in the determination of a directional vector g for which this condition holds, we can make use of the normalization constraint to comply with the desirable property of uniqueness of the technical efficiency value for all firms belonging to an output isoquant when projected onto the same benchmark isoquant, Lemma 2. It is possible to achieve these conditions imposing as normalizing constraint $wg = 1$, which results in yet another desirable by-product: The interpretation of the inefficiency value in terms of a monetary metric, as presented in Zofio et al. (2013). Indeed, it is easy to see that if the unitary normalization constraint

holds in (18), then cost, allocative and technical inefficiency are measured in monetary values.

Let $x^1 \in L(y^1) \cap \text{Isoq } L(y^2)$. Determining the value of the reversed directional distance function along with the directional vector g with the normalization constraint $wg = 1$ for x^1 implies solving the following non-linear optimization program:

$$\begin{aligned} \bar{D}^R(x^1, y^1, y^2, w) &= \underset{\beta, g}{\text{Max}} \beta \\ \text{s.t.} \\ y^1 &= f(x^1 - \beta g), \\ \frac{w(x^1 - \beta g) - C(y^1, w)}{wg} &= \frac{wx^1 - C(y^2, w)}{wg}, \\ wg &= 1, \\ \beta &\geq 0, \end{aligned} \quad (20)$$

where the first constraint ensures that the projected vector belongs to the observed output level—i.e., $\text{Isoq } L(y^1)$, the second one keeps constant the allocative inefficiency value measured at $\text{Isoq } L(y^2)$ on the firm's projection on $\text{Isoq } L(y^1)$, and the final constraint imposes the desired normalizing constraint that makes technical efficiency equal for all firms belonging to $\text{Isoq } L(y^2)$ and projected onto $\text{Isoq } L(y^1)$, while measuring cost inefficiency in monetary values.¹³

The next proposition states that technical inefficiency is the same for firms located on the same output isoquant, extending the property summarized in Lemma 2 to the non-homothetic case.

Proposition 4 *Let f be a production function. Let $x^1, x^2 \in \text{Isoq } L(y^2)$. Then, $\bar{D}^R(x^1, y^1, y^2, w) = \bar{D}^R(x^2, y^1, y^2, w)$.*

Proof By the second constraint in (20), $w(x^1 - \beta g) - C(y^1, w) = wx^1 - C(y^2, w)$, which is equivalent to $\beta = C(y^2, w) - C(y^1, w)$ since $w(x^1 - \beta g) = wx^1 - \beta wg = wx^1 - \beta$. In this way, $\bar{D}^R(x^1, y^1, y^2, w) = \bar{D}^R(x^2, y^1, y^2, w)$ for any $x^1, x^2 \in \text{Isoq } L(y^2)$. ■

Finally, we are able to prove that cost inefficiency can be decomposed in a reverse way without resorting to inconsistent second-order residual terms as $AAE(x^1, y^2, w)$ in Eq. (17).

Proposition 5. *Let f be a production function. Let $x^1 \in L(y^1) \cap \text{Isoq } L(y^2)$ and let (β^*, g^*) an optimal solution of (20). Then,*

$$\begin{aligned} CI(x^1, y^1, w; g^*) &= \frac{wx^1 - C(y^1, w)}{wg^*} \\ &= AI(x^1, y^2, w; g^*) + \bar{D}^R(x^1, y^1, y^2, w). \end{aligned}$$

¹³ Given the normalizing constraint, the second restriction can be simplified as follows: $C(y^2, w) - C(y^1, w) = \beta$. We thank one of the referees for this suggestion.

Proof $\bar{D}^R(x^1, y^1, y^2, w) = \beta^*$, which is equal to $C(y^2, w) - C(y^1, w)$ by the first part of the proof of Proposition 4. On the other hand, $AI(x^1, y^2, w; g^*) = (wx^1 - C(y^2, w))/(wg^*)$ by (19), which coincides with $wx^1 - C(y^2, w)$ since $wg^* = 1$. Finally, summing $AI(x^1, y^2, w; g^*)$ and $\bar{D}^R(x^1, y^1, y^2, w)$ we get $wx^1 - C(y^2, w) + C(y^2, w) - C(y^1, w) = wx^1 - C(y^1, w)$, which is equivalent to $\frac{wx^1 - C(y^1, w)}{wg^*} = CI(x^1, y^1, w; g^*)$.

3.3 Examples with non-homothetic Cobb-Douglas production functions: NH-CD

Example 1 We now retake our non-homothetic Cobb-Douglas technology $2^{y-1} = x_1^{0.5} x_2^{0.5y}$ illustrated in Fig. 1 and solve problem (20) for a set of firms producing different output levels. Results are reported in Table 2. Considering once again $(x^1, y^1) = (x_1^1, x_2^1, y^1) = (4, 1, 1)$ as the evaluated firm, and the optimal cost minimizing input quantities for the observed output level $y^1 = 1$, $(x^*, y^1) = (x_1^*, x_2^*, y^1) = (1, 1, 1)$, we firstly solve model (20), determining in this way technical inefficiency $\bar{D}^R(x^1, y^1, y^2, w) = 1\$$ and $g^* = (g_{x_1}^*, g_{x_2}^*) = (0.2679, 0.7321)$. Additionally, to obtain a value for the allocative inefficiency of this firm, we resort to the final expression in (19), with $AI(x^1, y^2, w; g^*) = \$2$. Finally, by Proposition 5, $CI(x^1, y^1, w; g^*) = \$2 + \$1 = \$3$. Using model (20) to decompose cost inefficiency ensures that solving technical inefficiency keeps allocative inefficiency constant when measured at $L^{NH}(y^1 = 1)$ and $L^{NH}(y^2 = 2)$. Moreover $\bar{D}^R(x^1, y^1, y^2, w) = \1 representing technical inefficiency is measured in the reference direction $g^* = (g_{x_1}^*, g_{x_2}^*) = (0.2679, 0.7321)$, with the projected technical efficient input quantities on Isoq $L(y^1)$ corresponding to $\hat{x}^1 = (x_1^1 - \bar{D}^R(x^1, y^1, y^2, w)g_{x_1}^*, x_2^1 - \bar{D}^R(x^1, y^1, y^2, w)g_{x_2}^*) = (3.7321, 0.2679)$.¹⁴

Table 2 presents the same analysis for two more firms, $(x_1^2, x_2^2, y^1) = (1, 2, 1)$ and $(x_1^3, x_2^3, y^1) = (0.25, 4, 1)$, once again targeting output level $y^1 = 1$ but employing more inputs than those required to be technically efficient, or targeting $y^2 = 2$ but yielding $y^1 = 1$ only. While both firms exhibit equal technical inefficiency $\bar{D}^R(x^2, y^1, y^2, w) = \bar{D}^R(x^3, y^1, y^2, w) = \1 , by belonging to the same isoquant $L^{NH}(y^2 = 2)$ and taking as reference to measure technical efficiency $L^{NH}(y^1 = 1)$, $(x_1^2, x_2^2, y^1) = (1, 2, 1)$ is allocative efficient at $L^{NH}(y^2 = 2)$, which is not the case for $(x_1^3, x_2^3, y^1) = (1, 4, 1)$. Taking advantage of Fig. 1, we confirm that the second firm belongs to the expansion path, thereby satisfying $x_2^* = (w_1/w_2)x_1^*y$. This is not the case for the third firm with $AI(x^3, y^3, w; g) = 1.125 > 0$. Table 2 also presents results for two more sets of firms targeting larger

outputs $L^{NH}(y^3 = 3)$ and $L^{NH}(y^4 = 4)$. All firms present the same technical inefficiency scores, either 0.7534 or 0.6635, with their associated individual directional values g , and different allocative efficiency values so as to totalize the observed cost inefficiency. As intended, the generalized approach yields a unique technical inefficiency value for all these firms, while ensuring that their allocative efficiency is kept constant.

Example 2 Building upon the previous *H-CD* example by Kopp and Diewert (1982), we modify it so as to observe the non-homothetic specification presented in Eq. (14) with $h(y) = 4468.507y^{-1}$ and $c(y) = 0.7y$, resulting in the following function: $4,468.507y^{-1} = 0.049x_1^{0.25}x_2^{0.1}x_3^{0.7y}$.¹⁵ This formulation ensures that the original input vector under evaluation $(x^1, y^1) = (x_1^1, x_2^1, x_3^1, y^1) = (500, 50, 162754, 1300)$ yields the intended but unrealized output target $y^2 = 1523.72$, as in the *H-CD* case. Table 3 presents the results for three sets of firms targeting successive output levels $L^{NH}(y^2 = 1523.72)$, $L^{NH}(y^3 = 1750)$ and $L^{NH}(y^4 = 2000)$, while the reference isoquant for the standard decomposition corresponds to Isoq $L(y^1 = 1300)$. Solving Eq. (20) for the firms targeting $L^{NH}(y^2 = 1523.72)$ results in the same value of technical inefficiency $\bar{D}^R(x^1, y^1, y^2, w) = \7423.93 with different directional vectors $g = (g_{x_1}, g_{x_2}, g_{x_3})$ —in the case of the first firm $g^1 = (g_{x_1}^1, g_{x_2}^1, g_{x_3}^1) = (0.0003, 0.0002, -0.0260)$. This firm presents a rather large individual allocative inefficiency $AI(x^1, y^2, w; g) = \$6,968,669.17$, with respect to the minimum cost of producing $y^2 = 1,523.72$, which is in sharp contrast to the second firm $(x_1^2, x_2^2, x_3^2, y^1) = (0.05, 0.05, 16,3211.40, 1300)$ that is the one allocative efficient at Isoq $L(y^2 = 1523.72)$ —note how the non-homothetic specification results in a drastic reduction of the optimal input demands for the first two inputs, along with a huge drop in the difference between minimum production costs for alternative isoquants, as they that get much closer together in the input space when compared to the homothetic specification. A similar scheme is obtained for the two additional sets of three firms producing $L^{NH}(y^3 = 1750)$ and $L^{NH}(y^4 = 2000)$.

4 Conclusions

The decomposition of cost efficiency *a la* Farrell in a parametric context, following the seminal work by Kopp and Diewert (1982) and subsequent refinements, for the case of non-homothetic technologies, calls for its

¹⁴ Note that for some firms it is necessary to increase the quantity of some inputs to realize their technically efficient amount; a result that stems from the flexibility of the directional distance function, and that could not be achieved with radial measures.

¹⁵ Kopp and Diewert (1982) further illustrate their model with an additional four-input non homothetic, non-neutral technical change translog frontier cost function, which also results in an inconsistent decomposition of cost inefficiency under the new approach, since their method keeps inputs' proportions (mix) fixed, as required by the radial approach.

Table 2 Reversed cost efficiency decomposition with a *NH-CD* technology: $2^{y-1} = x_1^{0.5} x_2^{0.5y}$

	Input x_1	Input x_2	Cost inefficiency (CI)	Allocative inefficiency (AI)	Technical inefficiency (TI)	gx_1	gx_2	\hat{x}_1	\hat{x}_2
$L^{NH}(y^2 = 2)$ $C(y^2, w) = 3$	4.0000	1.0000	3.0000	2.0000	1.0000	0.2679	0.7321	3.7321	0.2679
	1.0000	2.0000	1.0000	0.0000	1.0000	0.0000	1.000	1.0000	1.0000
	0.2500	4.0000	2.2500	1.2500	1.0000	−0.0941	1.0941	0.3441	2.9059
$L^{NH}(y^3 = 3)$ $C(y^3, w) = 3.5095$	4.0000	1.5874	3.5874	2.0779	1.5095	0.1220	0.8780	3.8158	0.2621
	0.8774	2.6321	1.5095	0.0000	1.5095	−0.0812	1.0812	1.0000	1.0000
	0.2500	4.0000	2.25	0.7404	1.5095	−0.1215	1.1215	0.4334	2.3070
$L^{NH}(y^4 = 4)$ $C(y^4, w) = 3.7893$	4.0000	2.0000	4.0000	2.2107	1.7890	0.0234	0.9766	3.9581	0.2526
	0.7579	3.0314	1.7893	0.0000	1.7890	−0.1353	1.1353	1.0000	1.0000
	0.2500	4.0000	2.2500	0.4607	1.7890	−0.1473	1.1473	0.2129	3.9629

Selected firms targeting different output levels larger than $L(y^1 = 1)$, $C(y^1, w) = 2$

reexamination in the light of recent developments in economic theory—Bogetoft et al. (2006), Aparicio et al. (2015). Contrarily to the case of homothetic technologies, defining technical efficiency by way of the radial input distance function is incorrect if the actual output level targeted by the firm, constituting the reference for allocative efficiency measurement, does not correspond to that observed—as in the standard approach. This is because the radial contraction of inputs resulting in cost savings cannot be surely ascribed to technical improvements, as demand for inputs, which under non-homotheticity depends on the intended output levels, varies along the projection path, and therefore allocative efficiency—defined as the deviation of marginal rates of substitution from relative inputs prices—changes over the input production possibility set.

From the perspective of the cost efficiency decomposition, this result invalids the residual nature of allocative efficiency, which is brought to the forefront of the analysis requiring further attention. Taking as departing point the standard decomposition proposed by Farrell (1957), and relying on Bogetofts et al.'s (2006) results, we emphasize that for homothetic technologies the radial distance function characterizes technical efficiency in a consistent way by leaving the allocative inefficiency counterpart unchanged, and based on the same criterion, we show that for non-homothetic technologies an equivalent definition of technical efficiency can be proposed in terms of the directional distance function.

The consistency of the new approach is based on the so called reversed cost efficiency decomposition, which is equivalent to the Farrell approach in the homothetic case, but yields alternative technical and allocative values when

non-homothetic technologies are involved. We prove several results that are needed to implement the new methodology, ensuring that once a given output level is considered for the reverse approach, allocative efficiency is first determined, and its technical efficiency counterpart rather than being still associated to the radial input measure, can be rightly determined by calculating the directional vector associated to the directional distance function. The new decomposition for non-homothetic technologies mirrors the properties of the standard approach, including that the technical efficiency for firms situating in a given isoquant is the same when projected onto the same reference isoquant, thanks to a suitable normalization condition. An additional property of our choice of normalization constraint is that inefficiencies are measured in monetary values.

We introduce the necessary optimization programs that are needed to implement the model empirically in a parametric setting, allowing us to calculate the directional distance function and its associated directional vectors, and illustrate the new methodology with Cobb-Douglas specifications. However, our theoretical model is presented in general terms, and therefore can be implemented also within semi-parametric or non-parametric frameworks; in this latter case using mathematical programming techniques. Since the approximations of the production technology that these techniques yield are generally non-homothetic, our results are as relevant for practitioners as those presented for the parametric case. They should be kept in mind by researchers concerned with the homotheticity properties of their production and distance function specifications, so as to correctly decompose overall efficiency into its technical and allocative components.

Table 3 Reversed cost inefficiency decomposition with a NH -CD technology: $4,468.507^{y-1} = 0.049,0^{0.25}, 0.1^{0.75} x_2^{-1} x_3^{0.75}$

	Input x_1	Input x_2	Input x_3	Cost inefficiency (CI)	Allocative inefficiency (AI)	Technical inefficiency (TI)	g^{x_1}	g^{x_2}	g^{x_3}	\hat{x}_1	\hat{x}_2	\hat{x}_3
$L^{NH}(y^2) = 1,523.72$ $C(y^2, w) = 10,610.341$	500.00	50.00	162,754.00	6,976,096.10	6,968,669.17	7426.93	0.0003	-0.0002	0.0260	497.75	51.61	162,561.16
	0.05	0.05	163,211.40	7426.93	0.00	7426.93	0.0000	0.0000	0.0154	0.05	0.05	163,097.14
	338.00	81.12	162,761.55	5,654,888.36	5,647,461.43	7426.93	0.0089	-0.0053	0.0254	272.23	120.82	162,572.92
$L^{NH}(y^3) = 1,750.00$ $C(y^3, w) = 10,615.924$	385.64	92.55	162,899.20	6,464,195.35	6,451,185.07	13,010.28	0.0108	-0.0065	0.0252	245.44	177.09	162,570.71
	0.05	0.05	163,297.30	13,010.28	0.00	13,010.28	0.0000	0.0000	0.0154	0.05	0.05	163,097.14
	817.08	22.86	162,892.83	10,247,353.65	10,234,343.36	1,3010.28	0.0110	-0.0067	0.0276	673.56	109.49	162,534.22
$L^{NH}(y^4) = 2,000.00$ $C(y^4, w) = 10,620.626$	647.84	64.78	163,010.15	9,062,498.73	9,044,786.37	17,712.36	0.0113	-0.0068	0.0264	447.24	185.77	162,543.06
	0.05	0.05	163,369.63	17,712.36	0.00	17,712.36	0.0000	0.0000	0.0154	0.05	0.05	163,097.14
	927.89	25.96	163,010.34	11,646,681.31	11,628,968.96	17,712.36	0.0113	-0.0068	0.0274	727.33	146.99	162,525.53

Selected firms targeting different output levels larger than $L(y^1) = 1300$, $C(y^1, w) = 10,602,913.90$, KD (1982)

Acknowledgements We would like to express our gratitude to two anonymous referees for their helpful comments, as well as to the Spanish Ministry for Economy and Competitiveness for supporting this research under grant MTM2013-43903-P.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Aigner DJT, Chu SF (1968) On estimating the industry production function. *Am Econ Rev* 58(4):826–839
- Aparicio J, Pastor JT, Zofío JL (2015) How to properly decompose economic efficiency using technical and allocative criteria with non-homothetic DEA technologies. *Eur J Oper Res* 240:882–891
- Aparicio J, Pastor JT, Zofío JL (2017) Can Farrell's allocative efficiency be generalized by the directional distance function approach? *Eur J Oper Res* 257:345–351
- Blackorby C, Diewert WE (1979) Expenditure functions, local duality, and second order approximations. *Econometrica* 47(3):579–602
- Bogetoft P, Färe R, Obel B (2006) Allocative efficiency of technically inefficient production units. *Eur J Oper Res* 168:450–462
- Boussemart J-P, Briec W, Leleu H (2010) Linear programming solutions and distance functions under α -returns to scale. *J Oper Res Soc* 61(8):1297–1301
- Boussemart J-P, Briec W, Peypoch N, Tavera C (2009) α -returns to scale in multi-output technologies. *Eur J Oper Res* 197(1):332–339
- Chambers RG, Chung Y, Färe R (1996) Benefit and distance functions. *J Econ Theory* 70:407–419
- Chambers RG, Chung Y, Färe R (1998) Profit, directional distance functions, and nerlovian efficiency. *J Optim Theory Appl* 98(2):351–364
- Chambers RG, Mitchell T (2001) Homotheticity and non-radial changes. *J Prod Anal* 15:31–39
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2/6:429–444
- Christensen LR, Greene WH (1976) Economies of scale in U.S. electric power generation. *J Polit Econ* 84(4):655–676
- Debreu G (1951) The coefficient of resource utilization. *Econometrica* 19(3):273–292
- Delis M, Iosifidi M, Tsionas EG (2014) On the estimation of marginal cost. *Oper Res* 62(3):543–556
- Färe R (1988) *Fundamentals of production theory*. Springer, New York
- Färe R, Grosskopf S (1997) Profit efficiency, farrell decompositions and the mahler inequality. *Econ Lett* 57(3):283–287
- Färe R, Lovell CAK (1978) Measuring the technical efficiency of production. *J Econ Theory* 19:150–162
- Färe R, Mitchell T (1993) Multiple outputs and homotheticity. *South Econ J* 60(2):287–296
- Färe R, Primont D (1995) *Multi-output production and duality: theory and applications*. Kluwer Academic, Dordrecht
- Farrell MJ (1957) The measurement of productive efficiency. *J R Stat Soc Ser A* 120:253–281
- Jacobsen SE (1970) Production correspondences. *Econometrica* 38(5):754–768
- Kopp RJ (1981) The measurement of productive efficiency: a reconsideration. *Quart J Econ* 96:477–503

- Kopp RJ, Diewert WE (1982) The decomposition of frontier cost function deviations into measures of technical and allocative efficiency. *J Econ* 19:319–331
- Kumbhakar SC (1997) Modeling allocative efficiency in a translog cost function and cost share equations: an exact relationship. *J Econ* 76(1/2):351–356
- Kumbhakar SC, Lovell CK (2000) Stochastic frontier analysis. Cambridge university press, New York, USA
- Luenberger DG (1992) New optimality principles for economic efficiency and equilibrium. *J Optim Theory Appl* 75(2):221–264
- Madden P (1986) Concavity and optimization in microeconomics. Blackwell Publishers, New York, NY
- Mangasarian OL (1994) Nonlinear programming. Vol. 10 of classics in applied mathematics. SIAM, Philadelphia, PA
- Mensah YM (1994) A simplification of the Kopp-Diewert method of decomposing cost efficiency and some implications. *J Econ* 60:113–144
- Sato R (1975) The more general class of CES functions. *Econometrica* 43:999–1003
- Sato R (1977) Homothetic and non-homothetic CES Production Functions. *Am Econ Rev* 67(4):559–559
- Schmidt P, Lovell C (1979) Estimating technical and allocative inefficiency relative to stochastic production and cost frontiers. *J Econ* 9(3):343–366
- Shephard RW (1953) Cost and production functions. Princeton University Press, Princeton, NJ
- Shephard RW (1970) Theory of cost and production functions. Princeton University Press, Princeton, NJ
- Silberberg E, Suen W (2000) The structure of economics: a mathematical approach. McGraw-Hill, New York
- Simar L, Vanhems A, Wilson PW (2012) Statistical inference for DEA estimators of directional distances. *Eur J Oper Res* 220:853–864
- Zieschang KD (1983) A note on the decomposition of cost efficiency into technical and allocative components. *J Econ* 23(3):401–405
- Zofio JL, Pastor JT, Aparicio J (2013) The directional profit efficiency measure: on why profit inefficiency is either technical or allocative. *J Prod Anal* 40(3):257–266