SUSTAINABLE EFFICIENCY OF FIRMS WHEN NEW SUSTAINABILITY TARGETS ARE INTRODUCED

Mondelaers[i], K., Kuosmanen[ii], T., Van Passel[iii], S., Buysse[iv], J., Lauwers[i], L and Van Huylenbroeck[iv], G.

[i]Institute for Agricultural and Fisheries Research, Social Sciences Unit, Belgium

[ii]Helsinki School of Economics, Finland

[iii]Hasselt University, Belgium

[iv]Ghent University, Belgium

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

Abstract
There is a high potential for simultaneously increasing sustainability of the earth system and economic development by removing inefficiencies currently present both at the production input and output side. In this paper a static view on sustainability is employed, by introducing capacity constraints as the boundaries above (or below) which the system cannot maintain its stable state. Currently these capacity constraints are often not respected. In this paper it is shown how the efficiency improvement pathway of an industry and the firms within it can be calculated to come to a sustainable, profit maximizing state, given the existence of these capacity constraints.

Key-words
Efficiency analysis, sustainability, DEA, directional distance vectors

1. Introduction

The term 'sustainable development' has since its first introduction to the general public after Rio[1], 1992, gained considerable attention. Many contributions have been made for the assessment of sustainable development, some from fragmentary, others from more holistic nature. The term sustainable development seems in itself contradictory: 'to sustain' versus 'to develop'. This contradiction in fact says that development is subject to the constraint of guaranteeing the sustainability of some elements upon which the development is realized.

According to system theory, we can define a hierarchical system in which we find a supersystem and multiple subsystems, the latter being embedded in the former. The supersystem could be the earth system, while the subsystem could be a firm. Our final goal is to guarantee (or move to) a sustainable and welfare maximizing supersystem, with 'sustainable' defined as the ability of a system to continue over a certain time span (Hansen and Jones, 1996). What causes continuity (or failure) of the supersystem, can be related to the capacity constraints of the system, the thresholds above or below which the system cannot recover. A practical example is global warming, or at a more local level, the pollution of aquifers. The current level of resource consumption is not always sustainable (see IPCC, 2007 f.e.), although this is now assumed in most economic models. To overcome this bottleneck, we could start from the carrying capacity constraints of the supersystem.

Voinov and Farley (2007), who approach the concept of sustainability from a system's perspective, argue that our current obsession with sustaining a growth-driven economic system probably comes at the expense of the ecologic supersystem. Advances in resource efficiency can be overcompensated because higher efficiency may lead to increased use of (environmental) resources. This is called the rebound effect (Mayumi et al., 1998, Herring and Roy, 2002). Voinov and Farley (2007) give the example of cancerous cells that can be successful as a subsystem and simultaneously have a pernicious effect on the human body as a supersystem. Because continuity of the supersystem and subsystem are not considered over the same time span, this conflicting situation can occur as the subsystem does not experience an incentive to restrict its resource use (its longevity is much shorter than this of the supersystem).

In this paper we focus on the efficiency score of firms in an industry when this industry is characterized by overconsumption of scarce resources and/or overproduction of bad outputs. There is a potential for meeting industry sustainability targets by removing (part of) the firm level inefficiency. Moreover, there might even be a possibility to create more firm level output and profit, given the reaching of these industry constraints, when the remainder of inefficiency is removed. The objective of this paper is to obtain a single firm index of 'sustainable profit efficiency', a term which simultaneously captures the meeting of (higher level) sustainability targets and the creation of more firm level value (output or profit). This paper contributes to the literature on efficiency measurement by linking efficiency to absolute sustainability targets at macro level, a topic which is not yet deeply investigated. Several authors have associated inefficiency with firm level sustainability targets (such as Reinhard et al., 1999, Fare et al., 2004a, Coelli et al., 2007). Others have focused on the determination of the industry's efficiency score based upon the firms' inefficiencies (Blackorby and Russell,1999, Li and Ng, 1995, Fare and Zelenyuk, 2003), while yet another stream of efficiency literature focuses on nonradial efficiency scores, the so called directional distance functions, with main contributions from Chambers et al., 1998, Chung et al., 1997, based

[1]United Nations Conference on Environment and Development (UNCED), Rio de Janeiro, 3-14 June 1992

upon the work of Luenberger, 1992 and recent advancements by Briec et al. (2003) and Kuosmanen et al. (2009). We will use elements from these three streams of literature in efficiency analysis to link inefficiency with sustainability targets. The focus on macro sustainability targets is of concern, as policy makers and practitioners are seeking the optimal way to distribute the burden of production over participants.

The paper unfolds as follows. In a first part the sustainability targets are mathematically defined. The second part focuses on the potential interplay between efficiency and sustainability and identifies different states of the industry. In the subsequent part 4 we focus on the most common of these states, an inefficient and unsustainable industry, and propose a way to calculate the sustainable profit efficiency of the firms in this industry. The method is then illustrated by means of the Belgian dairy farming case. We end with conclusions.

2. **Defining capacity constraints**

In ecology carrying capacity is defined as the maximal population size of a given species that an area can support without reducing its ability to support the same species in the future (Daily and Ehrlich, 1992). Seidl and Tisell (1999) mention that a judgement of an environmental situation or the decision of limits (e.g. the carrying capacity) is influenced by value-judgements and institutional settings. They also extend the biophysical concept with a social counterpart. Social carrying capacity, shaped by human consumption patterns, technologies, infrastructure and so on, refers to the maximal population size that could be sustained under various social systems, while biophysical carrying capacity reflects the maximal population size that can be sustained biophysically under given technological capabilities. The latter can only be higher or equal to the former. In this text we define a carrying capacity constraint as the total amount of a resource that can be consumed without endangering the sustainability of the surrounding biophysical ecosystem. Daily and Ehrlich (1992) provides a classification of resources based upon key sustainability issues. They distinguish between resources that are not necessarily degraded or dispersed in use while providing free services to humanity, such as microbial nutrient cyclers, natural pest control agents and pollinators (mutual benefit), opposite to resources such as food, drinking water and energy that are necessarily consumed, dispersed or degraded to derive benefits from them. The second distinction they make relates to renewable versus non renewable resources, with the latter being mainly flow limited while the former are generally stock limited. Last distinction they make is between availability of substitutes or not. Resources, such as fresh water, biodiversity and fertile soils, which have no substitute, are termed essential. To go from this classification to a capacity constraint, an additional concept is needed. Daily and Ehrlich (1992) therefore defined a ‘maximum sustainable use’ (MSU) of a resource depending on its classification. For degradable, substitutable resources Daily and Ehrlich (1992) define a quasi-sustainable consumption rate equivalent to the rate of generation of the substitutes. For a renewable essential resource that is necessarily consumed, degraded or dispersed in the extraction of value from it, the MSU is equivalent to its renewal rate.

The authors also distinguish between scarce inputs and undesired outputs. For the latter they define the maximum sustainable level of abuse (MSA), or the maximal sustainable emission rate of a pollutant into the environment. This is the level of emission that produces the highest concentration of pollutant that can be tolerated by the most sensitive system element.

In this paper we do not focus on the way the MSU and MSA are calculated, as this is resource and situation dependent. We assume that both the actual resource use and the MSU (or MSA) are known. In our reasoning, the MSU is synonymous to a constrained resource use.

The current resource consumption at a supra farm (f.e. regional) systemic level can be computed easily by aggregating over the different farms that are part of this systemic level. By means of environmental scientific studies we can obtain an estimate of the desirable resource consumption from a sustainability perspective. Imagine that this desired use (the MSU) is a percentage α of the current use. The supra farm level constraint can then be defined as:

$$\alpha \cdot \sum_k^K x_{1,k} \geq \sum_k^K x_{1,k}^c \quad (1)$$

with $x_{1,i}$ the actual use of input x_1 by firm i, $(1-\alpha)$ the desired (macro level) reduction in input use, K the total number of firms subject to this constraint. x^c indicates the constrained input use of firm k of input x_1. If not sufficient, input can be contracted further by reducing the output.

The case of a single resource constraint can be easily extended to multiple constraints. In this paper we only develop the reasoning for a single higher level resource constraint, due to space limitations.

3. **The interplay between efficiency and sustainability**

Following Figure 1, representing an industry's production function with resource (set) X and output Y, can be used to explain how sustainability and efficiency relate. To the left of the capacity constraint (the MSU), resource consumption is sustainable. The difference between inefficient resource use and unsustainable resource use is also depicted in Figure 1. Unsustainable resource use refers to the situation in which constant (natural) capital stocks are no longer maintained, or, put differently, that strong sustainability thresholds are surpassed. There is however no link with the output achieved. Inefficient resource use on its turn reflects a suboptimal relation between amount of resource used and output reached. We can distinguish between 5 cases, as shown in Figure 1.

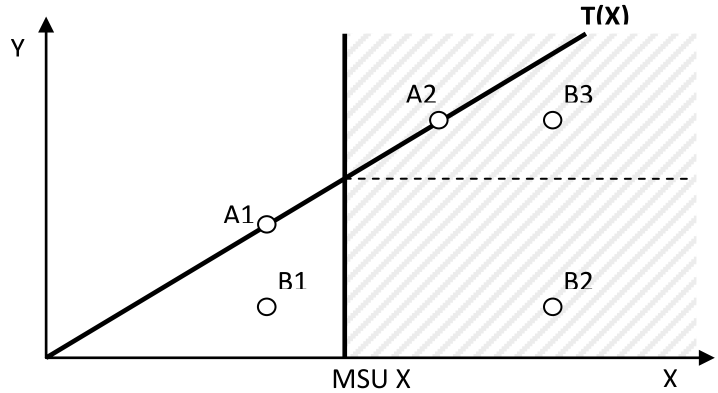

Figure 1. Sustainable versus efficient resource use of an industry. T(X) reflects the production frontier and MSU X the maximum sustainable use of X. Following categories can be distinguished: A1: efficient and sustainable resource use; A2: efficient but unsustainable; B1: inefficient but sustainable; B2: inefficient and unsustainable, but removing inefficiency suffices to achieve sustainability; B3: inefficient and unsustainable, and efficiency improvement does not suffice to achieve sustainability

Depending on whether the production frontier is situated within the capacity constraint or not, the current technology is suited for sustainable production. The value of the classification is that relative efficiency and absolute sustainability are not competing but complementary criteria. Knowing the status of the system according to this classification is important for making appropriate policy decisions. Cases A1 and B1 are not problematic as far as sustainability is concerned. In case B2 it is possible to maintain the current level of output by improving efficiency of resource use. Cases A2 and B3 require downscaling of output level; in B3 this can be complemented with efficiency improvement.

Figure 1 reflects the industry composed of different economic entities. The possible positions of these entities in the figure can also be grouped in these 5 categories. It is thus very well possible that an industry is in sustainable equilibrium, with some firms breaching their proportional share of the MSU, as this is compensated by other firms who remain below their proportional MSU.

In line with principles of sustainable development, we aim to maximize an industry's value creation, while simultaneously meeting the capacity constraints. To this end following possibilities are available: we can reduce the technical inefficiency in the industry, we can change the industry's and hence the firms' input mix, to reach the capacity constraints at the lowest (shadow) cost possible, and we can change the output generated. By decomposing the aggregate 'sustainable inefficiency' into different parts, we can yield more information on which components of inefficiency are present both at industry and at firm level. The decomposition at industry level is important for policy making, while the decomposition at firm (versus industry) level reveals information with respect to the competitive positioning of firms in this industry.

4. Unsustainable and inefficient resource use[2]

In figure 1, option A1 and B1 do not pose sustainability concerns. In both cases output can still be increased up to the point that the capacity constraint is met. In the case of A1, industry output can be expanded by consuming extra resources under an efficient technology regime. In the case of B1, this can be complemented with some efficiency improvement. B2 offers an interesting starting point for our discussion, because it is the most straightforward case (with exception of A1 and B1). In this case the economy is characterized by unsustainable and inefficient resource use, and removing inefficiency is sufficient to attain the sustainability constraint. The question is here how the sustainability constraint should be divided over the different firms.

In this paper it is the aim to obtain an index for a firm's sustainable profit inefficiency. In the static view employed in this paper, a firm is considered to be **sustainable profit efficient (SPE)** when it creates maximum economic value while sustainability constraints (at supra firm level) are met. We will distinguish between **sustainable technical efficiency (STE)**, **sustainable allocative efficiency (SAE)** and **sustainable profit efficiency**, which is a combination of both other efficiency measures.

As depicted in Figure 2 below, an industry B2 can remove technical input inefficiency until its members jointly meet the capacity constraint. The remaining inefficiency can be removed to maximize output under the constrained regime. The removal of a certain amount of input and output inefficiency can be projected on a single vector (IE+OE in Figure 2). There is a stream of literature focusing on nonradial efficiency measures and directional distance vectors (see Chambers et al., 1996, Chung, et al., 1997, Chambers et al., 1998 and more recently Lee et al, 2002, Fare et al., 2004, McMullen et al., 2007, Murty et al., 2007, Briec and Kerstens, 2009, Kuosmanen et al., 2009), which is helpful for our case, as we want to define the necessary input contractions for each firm in the industry to jointly meet the capacity constraint as well as the possible output increases for these firms. A directional distance vector seems appropriate then.

[2] Given the page limitation, only this case is treated, as it is the most common

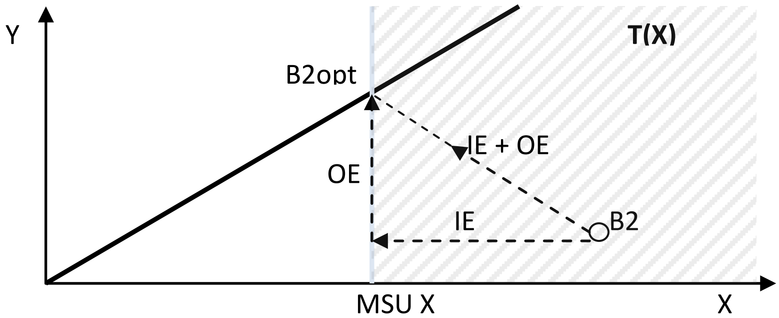

Figure 2. Input (IE) and output inefficiency (OE) removal potential for industry B2

In accordance to Fare et al. (2004), let $x \in R_+^N$ denote a vector of inputs and $y \in R_+^M$ a vector of outputs and the physical component of a technology denoted by *T*, where

$$T = \{(x, y): x \text{ can produce } y\} \quad (4)$$

T is assumed to be closed and convex with freely disposable inputs and outputs. The assumption of freely disposable outputs can be relaxed when bad outputs are present. Fare et al. (2004), in pursuit of Chambers et al. (1998), define the directional distance function on T as

$$\overrightarrow{D_T}(x, y; g_x, g_y) = max\{\beta: (x - \beta g_x, y + \beta g_y) \in T\} \quad (5)$$

where the nonzero vector $(g_x, g_y) \in R_+^N \times R_+^M$ determines the direction in which inputs are contracted and outputs are expanded.

Defining the appropriate directional distance vector is straightforward in the absence of sustainability constraints, because this boils down to measuring the output inefficiency, i.e. $g_x = 0$ and $g_y = y$. When sustainability constraints are present, we need a mechanism to define the optimal mix between g_x and g_y to ensure that output is maximized under the restriction of the capacity constraint. This will prove especially challenging for supra firm sustainability constraints, as we then need to find the optimal way of dividing the constrained input over the firms.

We can identify two types of directional distance vectors that are appropriate in our case:

1. a single directional distance vector (g_x, g_y) for all the firms in the industry
2. a directional distance vector ($g_{x,n}$, $g_{y,n}$) which is firm specific

Option 1, depicted in Figure 3 (full grey line), is appealing as it shows the general direction in which all the firms in an industry have to move to meet the capacity constraint and simultaneously improve output. It adequately solves the reallocation problem[3] which occurs when higher level capacity constraints need to be distributed over multiple firms. As all firms are evaluated in the same direction, the metric found can be easily interpreted. The single direction comes with a (minimal) cost, some potential industry output increase is lost compared to the case of the firm specific directional distance vector.

Option 2, the firm specific directional distance vector ($g_{x,n}$, $g_{y,n}$), as depicted in Figure 3 (dotted line), is also appealing as it allows to maximize the potential industry output given the capacity constraint. We could specify the x-component of the directional distance vector of firm n as $g_x X_n$, and the y-component as $g_y Y_n$. This means that the vector is dependent on a firm's x and y-level, but also that each vector shares a common part with the vectors of the other firms in the industry. This makes the reallocation mechanism of the constrained

[3] Imagine for example two identical firms and imagine that, given the capacity constraint, industry profit is maximized by reducing resource use of one of the two firms, keeping the use of the other constant. Although from an industry perspective it doesn't matter which firm's input use has to be reduced, from a firm perspective it does. The firm who's input use is reduced after optimization, receives a low efficiency score, opposite to the other. When the aim is to obtain a firm level inefficiency estimate, the firm's mirror should be the optimum from firm perspective, otherwise the micro level behavioral rule of private utility maximization is absent.

resource x over the firms still behaviorally interpretable. The more inefficient a firm is in the constrained resource x_c the more its directional distance vector projects it in the g_x direction. The more inefficient a firm is in the output direction, all else equal, the more it is projected in this direction. Firms inefficient in not-constrained resources are, logically, projected in the same direction as their counterpart efficient in these resources, as the directional vector only depends on x_c and y.

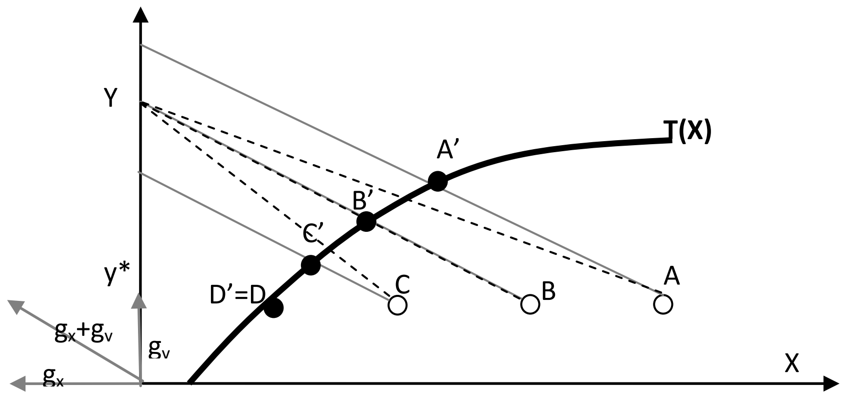

Figure 3. Grey line: improvement path for firms A, B, C and D in an industry when there is a single, industry wide directional vector (g_y; g_x). Dotted line: improvement path for firms in an industry when the directional vector (g_yY; g_xX) is partly firm specific.

4.1.1 Sustainable technical inefficiency

In this section the aim is to determine the output inefficiency of the firms in a industry, corrected for the industry's unsustainable input use. This correction is achieved by removing part of the firm's input inefficiency to the extent that the MSU is met. As can be seen in option B2 in Figure 1, removing only part of the input technical inefficiency at firm level is sufficient to guarantee that the joint input consumption of the firms in the industry meets the industry MSU. The industry does not need to reduce its input use further, as this level of input consumption already guarantees higher (eco)system sustainability[4]. The remaining inefficiency can be used to optimize firm level output. As such we obtain a measure for technical output inefficiency corrected for (higher level) sustainability targets, which indicates how much the output can be expanded given sustainability constraints.

When multiple outputs are present, we suggest to maintain the current output mix, as the decision maker has some (unobserved) reasons to choose this mix. In the section on sustainable allocative inefficiency the input and output mix is optimized to generate maximum firm profit in the presence of capacity constraints. As the real prices for the constrained resources are unknown and therefore need to be approximated, an indication of technical output inefficiency, even for multiple outputs, is informative.

The necessary input contraction *at industry level* is straightforward, as this is reflected by the relative difference between the capacity constraint and the current resource use. This could be used as starting point for the capacity constraint component (g_x) of the directional distance vector. This choice of the directional distance vector however does not guarantee that the capacity constraint is met, as some firms already operate on the efficiency frontier (esp. in a non parametric setting) and therefore they do not contribute to input reduction. Furthermore, each firm has potentially a different input inefficiency for the constrained (and unconstrained) resources, uses a different amount of the constrained input and has a different firm size.

[4] at least in a static setting

When a non parametric production model is used, the according piecewise linear production technology is formulated as (based upon Lee et al., 2002):

$$T(x) = \{(y): y \leq Y\lambda, x \geq X\lambda, e^T\lambda \leq 1, \lambda \in \mathcal{R}_+^i\} \quad (6)$$

The (nonlinear) nonparametric models below guarantee that industry output (as a sum of firm inefficiency corrected outputs) is maximized, while simultaneously meeting the sustainability target.

In the first model, the individual firm is projected on the efficiency frontier based upon vector ($g_x x$, $g_y y$). θ_l provides a direct measure of how far (x, y) must be projected along ($g_x x$, $g_y y$) to reach the frontier of $T(x)$ (Chambers et al., 1998). The directional vector in model 7 is firm specific, as it depends on x and y. Vector components g_x and g_y however are constant over the firms. The term ($1+g_y\theta_l$) indicates the remaining firm level output inefficiency after removal of a proportionate part of the input inefficiency ($1+g_x\theta_l$). As both indices relate to each other by means of θ_l, the technical output inefficiency can be used as a the efficiency measure. As there is only one direction that maximizes output given the capacity constraint, the constants g_x and g_y are not independent. Therefore, by fixing one, the complexity of the model below can be further reduced. If for example, g_x is set to -1 in the formula below, each firm's input $x_{c,l}$ is contracted with θ_l times $x_{c,l}$, while the output is expanded with g_y times θ_l times y_l. This approach assures that firms contribute to the capacity constraint according to their inefficiency.

$$Max\ Y = \sum_{l=1}^{L}\theta_l \quad (7)$$

$$\text{s.t. } \sum_{k=1}^{K}\lambda_{k,l}\, y_k \geq (1 + g_y\theta_l)y_l \qquad \text{for each k} \quad (7a)$$

$$\sum_{k=1}^{K}\lambda_{k,l}\, x_{k,c} \leq (1 + g_x\theta_l)x_{c,l} \quad (7b)$$

$$\sum_{k=1}^{K}\lambda_{k,l}\, x_{k,u} \leq x_{u,l} \quad (7c)$$

$$\sum_{l=1}^{L}(1 + g_x\theta_l)x_{c,l} \leq \alpha\sum_{l=1}^{L}x_{c,l} \quad (7d)$$

$$e^T\lambda \leq 1$$

$$\lambda, \theta \geq 0$$

with Y= industry output, l= firm index, k=also firm index, θ_l=firm distance parameter, g_y=output component of directional distance vector, g_x=capacity constraint component of directional distance vector, λ=weight of peer, y=output, x_c=constrained input, x_u=unconstrained input, e^T= matrix of 1's

By solving this model, the optimal vector and the firm directional distances are found. These distances, which reflect technical inefficiencies (Fare et al., 2004), indicate the maximal output achievable under the constrained regime. Based upon the right hand side of equations 7a and 7b, an output and capacity constraint related efficiency measure can be obtained, $(1 + g_y\theta)$ and $(1 + g_x\theta)$ respectively. As they are inextricably linked through vector (g_x, g_y), reporting the sustainable output technical inefficiency is sufficient as a measure for a firm's sustainable technical inefficiency.[5]

Whether option 1 or option 2 is chosen will not change the ordering of firms based upon their sustainable technical inefficiency. Option 1 yields a slightly higher overall output. There is a minor change in the obtained sustainable technical inefficiency scores, whereby firms which are most inefficient receive a better score in the model with firm specific directional distance vectors, at the cost of the more efficient. In the latter model, the difference in sustainable technical efficiency is thus less pronounced.

4.1.2 Sustainable allocative efficiency

The former procedure allows us to meet the capacity constraint and to optimize output generation at firm level. Given this, there is still potential to increase profit generation, by optimizing the input mix in function of input and output prices and as such removing the allocative inefficiency still present. Chambers et al. (1998) show the duality between the

[5] The model for the single industry wide directional distance vector is not reported due to space limitations

profit function and the directional technology distance function. We refer to their reading for a formal derivation. As shown in Figure 4, the inefficiency of a firm in an industry can be decomposed into a technical (TE) and an allocative part (AE, Chambers et al., 1998):

$$TE = \overrightarrow{D_T}(x, y; g_x, g_y) \quad (8)$$

$$AE = \frac{[\pi(p,w) - (py - wx)]}{(pg_y + wg_x)} - \overrightarrow{D_T}(x, y; g_x, g_y) \quad (9)$$

with $\pi(p, w)$ = maximal feasible profit; $py - wx$ = actual profit; $pg_y + wg_x$ = price normalization to avoid effect of proportional price changes; $\overrightarrow{D_T}(x, y; g_y, g_x)$ = directional distance function.

Specific for our case of MSU's is that the input price vector w (and potentially the output price vector) will change due to the introduction of a MSU. The point (x^*, y^*) is determined by the tangency between the frontier of *T(x)* and the line segment whose slope is given by an adjusted input price w^c.

The fact that the MSU is exceeded indicates that the constrained resource is not appropriately priced, i.e. the price does not reflect the real scarcity of the resource, or the real (external) cost, otherwise the industry would not be in this equilibrium. A possible explanation is given by Lawn (2007, p82), who argues that resource prices reflect reasonably well the relative scarcity of various resource types, but they do not adequately reflect the absolute scarcity of a resource type. The stock will determine the total inflow over time, but not the inflow at a particular point in time. Ecological economists argue that relative prices are mainly based upon flow based forces generated by the interacting market supply and demand forces. Oil is a good example. While oil stocks decrease over time, in particular periods prices can also decrease due to a drop in demand. We argue that the correct price is unknown due to absence of some information and/or the institutions to determine and enforce the exact price. This can both relate to absolute scarcity of a resource and external costs generated by exploiting a resource.

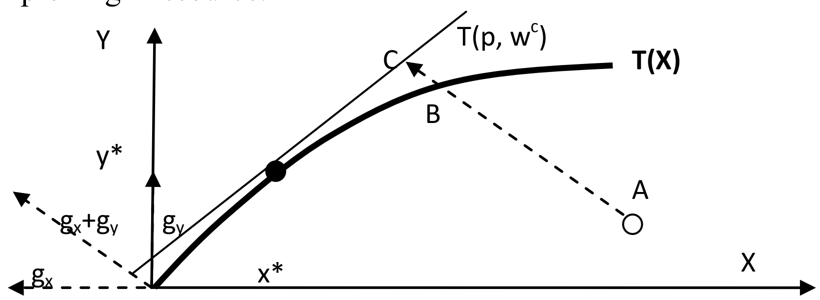

Figure 4. Technical and sustainable allocative efficiency measure in a directional distance framework (adapted from Chambers et al., 1998)

In an industry in which all firms are economically (i.e. technically and allocatively) efficient under an unconstrained regime (i.e without capacity constraint), the firms are (sustainable) allocative inefficient[6] under the constrained regime. With sustainable allocative inefficient we mean allocative inefficient when the MSU is respected. Introducing the MSU constraint will influence the supply and demand equilibrium and hence the prices. The industry's shadow price for the constrained resource can be used as a proxy for the real unknown market price, as suggested by Li and Ng (1995). Firms with a higher shadow price for this resource will be willing to pay for extra units of this resource while firms with a lower shadow price are interested to sell some units, until all firms reside in the new economic optimum.

Introducing the new sustainable price for the constrained resource will change the slope of the isoprofit line so that the capacity constraint at industry level is met. By defining the Kuhn-Tucker conditions of the industry profit function, we can obtain the shadow price

[6] In this example, every firm will have (the same) sustainable allocative inefficiency score.

which can serve as the new market price, see formula 12 and formula 13 further in the text. The dual of restriction 13g indicates the potential increase in market price for the constrained resource. This price can be used to reconstruct the new sustainable isoprofit line[7]. However, to apply this approach the industry technology has to be known. When both prices and technology are unknown, we can build further on the approach developed by Kuosmanen et al. (2009). To obtain shadow prices, they propose to minimize the sum of firms' profit inefficiencies, whereby the prices enter the inefficiency measure as decision variables of the optimization problem and the profit function is empirically determined. This industry distance function gives the lower bound for the true but unknown industry profit inefficiency:

$$IPE = \min_{w,p,\rho} \sum_{n\in v} \rho_n \tag{10}$$
$$\text{s.t. } \rho_n \geq (pY_m - wX_m) - (pY_n - wX_n) \qquad \forall m, n \in v$$
$$pg^y + wg^x = 1$$
$$p, w \geq 0$$

with IPE = estimated industry profit efficiency; w=unknown input price, p= unknown output price; ρ_n=profit inefficiency of firm n

When only one price is known, f.e. input price W_n, Kuosmanen et al. (2009) suggest to set input vector g_x equal to $1/W_n$ and all other elements of the directional vector equal to zero. We have no price information for the above problem, as all prices might vary given the introduction of the capacity constraint, but we know the industry directional distance vector (g_x, g_y) for the DD fixed at industry level, which we can use as normalization constraint in equation 10. This will allow us to determine the prices that minimize industry profit inefficiency. For the DD which varies at firm level, we can also isolate the fixed (g_x, g_y) component of the DD and use these for the normalization constraint (see formula 7). Given the obtained shadow prices, a firm's maximal profit relative to T can, in a nonparametric setting, be defined as:

$$\pi(p, w^c) = \max_{x,y,\lambda} p_c y - w_c x \tag{11}$$
$$\sum_{k=1}^{K} \lambda_k y_k \geq y_m \quad m = 1, \dots, M$$
$$\sum_{k=1}^{K} \lambda_k x_k \leq x_n \quad n = 1, \dots, N$$
$$\lambda_k \geq 0 \quad k = 1, \dots, K$$
$$\sum_{k=1}^{K} \lambda_k = 1$$

With the exception of the adapted prices p_c and w_c, this is the standard profit maximization model as also defined by Fare et al. (2004). Introducing the shadow prices p_c and w_c allows to calculate the expected maximum profit and the allocative inefficiency under the constrained regime, by applying formula 9.

Based upon the prior two sections, we can determine a firm's **sustainable allocative inefficiency** (SAE), its **sustainable technical inefficiency** (STE) and its **sustainable profit inefficiency** (SPE). Furthermore, with classic efficiency techniques, we can determine a firm's allocative inefficiency and technical inefficiency, in the absence of sustainability constraints. Comparison of both metrics will allow us to classify firms with respect to current resource use and sustainable resource use. It is possible that a firm is currently termed allocative inefficient, while it is at the same time sustainable allocative efficient.

5. Illustration: dairy farming under a single MSU

[7] For simplicity, we ignored the influence of introducing the MSU on the prices of the unconstrained inputs and the outputs (and hence assumed a partial equilibrium model). This could however be accommodated by introducing constraints for the other inputs, and allowing for a flexible output price.

The sample data used in this example consist of 271 Belgian dairy farms in 2004 and are derived from EU FADN. We assume that this sample is representative for the Belgian population of dairy producers. Physical output is measured as total milk production (in 1000 l) while inputs are concentrates (in ton), forage crops (in ha), labor use (in 100 hours), farm capital use (in 100€) and dairy cows (number). Some data conversions were necessary to construct these physical inputs from the costs reported in FADN. A stochastic frontier analysis revealed that these variables influence total milk production. We emphasize that this example is developed for an illustrative purpose.

For sake of simplicity, assume now that the global warming problem demands a 10% cut back in total number of (Belgian) dairy cows, as these are an important group of methane producers. The 90% remaining cows than acts as the capacity constraint or the MSU. Currently the sample livestock consists of 14.203 cows.

When all the radial output technical inefficiency would be removed, the total production of milk could be increased from 83.637.801 liter to 102.549.026 liter, which is a relative increase of 22,6%. When all input inefficiency is removed, the total number of cows drops back to 11.065, which is below the capacity constraint of 12782 cows, for a total output of 83.637.801 litre. By maintaining a stock of 12.782 cows (or 90% of the current stock), more output can be generated. The directional distance vector will help us to determine how much more. As explained in the text, we have two options for the directional distance (DD) vector: fixed at industry level (model 7) or firm specific (model 7bis). In this example we apply the firm specific DD vector.

For the model with firm specific DDs, the g_x and g_y component of the DD's are -0,4871 and 1, respectively. The total output generated by the industry under this scenario is 96.844.516 liter milk. Figure 5 below shows the obtained sustainable technical output inefficiencies (STE) under the MSU regime, for variable DD. The closer to unity, the more sustainable efficient the firm is.

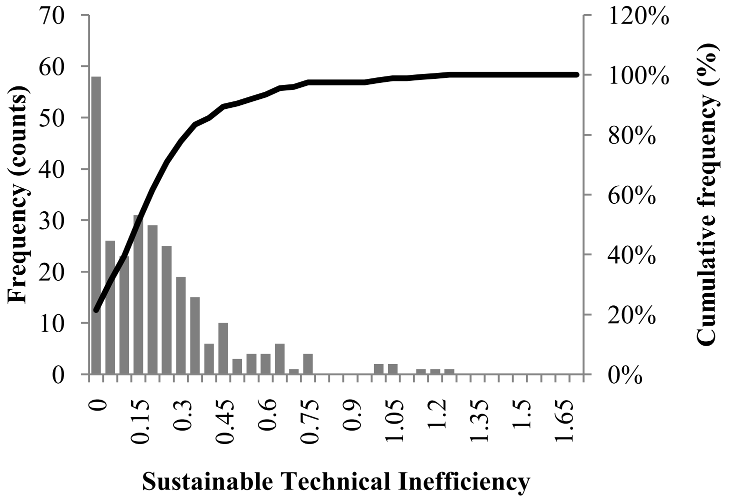

Figure 5. Histogram of Sustainable technical output inefficiency (STE) of the firms when the DD-vector varies over firms (STEv). Grey bars indicate count frequency, black line indicates cumulative frequency in %.

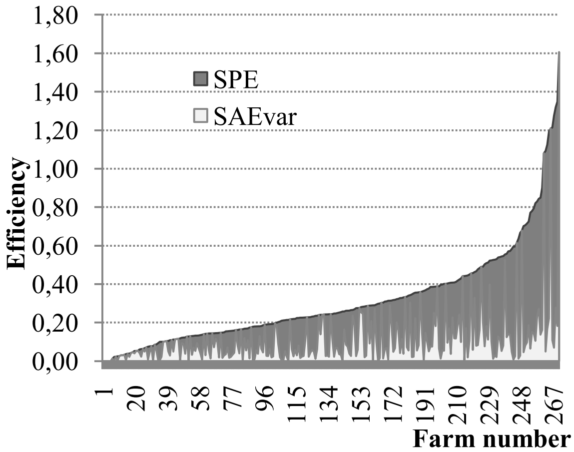

Figure 6. SPE and SAE-values for the firms in the sample, ranked in order of increasing SPE. The difference between SPE and SAE is the STE. The higher the SPE the lower the profit efficiency

With respect to sustainable allocative inefficiency, we first have to determine the industry shadow prices, by applying formula 10. The prices for labor, farm capital use, concentrates, forage crops, cows and milk might change due to the introduction of the MSU. We used normalization constraint $p+g_xw=1$, with p the shadow price for milk, w the shadow price for cows and g_x defined earlier in the example (-0.4871). Remember that g_y was arbitrarily set to 1. Shadow prices obtained are given in Table 1.

Based upon these prices we can now calculate the sustainable profit efficiency (SPE) and the sustainable allocative efficiency (SAE), according to formula 8, 9 and 11. The sustainable allocative inefficiency is simply obtained by subtracting STE from SPE. According to formula 9, and taking into account that our DD vectors vary per firm, a firm *i*'s sustainable profit efficiency is normalized by $pg_y y_i + wg_x x_i$, which gives it a straightforward interpretation. A SPE of 0.1 indicates that sustainable profit can be raised by 10% by improving sustainable technical and sustainable allocative inefficiency. The bars show the count frequencies (on a total of 271 farms), while the black line is the cumulative frequency percentage. The majority of farms has an SPE below 0.5, indicating that they can improve their sustainable profit up to 50% of the profit they now would attain under the sustainability scenario.

Figure 6 combines the firm SPE and SAE (and hence also the STE, as this is the difference between both), in one figure. The values are ranked in increasing order of SPE[8].

Table 1. Normalized shadow prices for the model with variable DD vector and average normalized shadow prices for the model with variable DD vector

	Milk output (in 1000l)	Cows	Labor (100h)	Capital (in 100€)	Concentrates (ton)	Fodder (in ha)
Shadow price variable DD	0.7544	0.5042	0.0138	0.0087	0.0192	0.0610

6. **Discussion and conclusion**

What can we learn from this decomposition of the firm's inefficiency into different components? The directional distance component indicates in which direction the firm should improve the management of its inputs in order to reduce the use of the constrained input to a sustainable level and to increase the output. The variable directional distance vector is extremely helpful to determine how much constrained input inefficiency should be removed to guarantee that the capacity constraint, and thus the sustainability target, is met. The remainder inefficiency can then be removed at the output side to increase output generation and as such increase the development side of 'sustainable development'. The sustainable allocative inefficiency on its turn indicates to what extend input substitution can increase the firm's profit.

The main question is however whether it is defendable to compare the firm with its peer on the frontier based upon the directional distance vector. There are different pathways for individual firms that still guarantee that the higher level capacity constraint is met. The main advantage of using the directional distance vector to determine a firm's reduction in constrained input use is that the decision rule is not arbitrarily imposed but based upon differences in efficiency between firms. From a sustainable development perspective, 100% efficiency guarantees that most value is created for a given input use. Therefore, more efficient firms are entitled to more of the constrained resource. The directional distance can be interpreted in a similar way as technical inefficiency. By improving the management of the resources, some constrained input and some output 'inefficiency' can be removed, keeping the non constrained inputs constant. The vectors shown in Figure 3 are a weighting between technical input and output efficiency. The weighting thereby depends on the size of the capacity constraint. If we want to steer the economy towards sustainable development, i.e. growth with sustainable resource use, it is helpful to have an overall direction. The directional distance vector gives this direction. From a sustainability perspective it is only relevant that the MSU is not exceeded. From a micro-economic perspective it might be interesting to know what the sustainable allocative inefficiency is, to maximize profit in monetary terms, ceteris paribus.

[8] An illustration of more than 1 constraint simultaneously can be obtained upon request

7. References

Blackorby, C. and Russell, R. R. (1999) "Aggregation of efficiency indices", Journal of Productivity Analysis, Vol 12 No 1, 5-20.

Briec, W. and Kerstens, K. (2009) "Infeasibility and Directional Distance Functions with Application to the Determinateness of the Luenberger Productivity Indicator", Journal of Optimization Theory and Applications, Vol 141 No 1, 55-73.

Briec, W., Dervaux, B., and Leleu, H. (2003) "Aggregation of directional distance functions and industrial efficiency", Journal of Economics-Zeitschrift Fur Nationalokonomie, Vol 79 No 3, 237-61.

Chambers, R. G., Chung, Y. H., and Fare, R. (1996) "Benefit and distance functions", Journal of Economic Theory, Vol 70 No 2, 407-19.

Chambers, R. G., Chung, Y., and Fare, R. (1998) "Profit, directional distance functions, and Nerlovian efficiency", Journal of Optimization Theory and Applications, Vol 98 No 2, 351-64.

Chung, Y. H., Fare, R., and Grosskopf, S. (1997) "Productivity and undesirable outputs: A directional distance function approach", Journal of Environmental Management, Vol 51 No 3, 229-40.

Coelli, T., Lauwers, L., and Van Huylenbroeck, G. (2007) "Environmental efficiency measurement and the materials balance condition", Journal of Productivity Analysis, Vol 28 No 1-2, 3-12.

Daily, G. C. and Ehrlich, P. R. (1992) "Population, sustainability, and earths carrying-capacity", Bioscience, Vol 42 No 10, 761-71.

Fare, R. and Zelenyuk, V. (2003) "On aggregate Farrell efficiencies", European Journal of Operational Research, Vol 146 No 3, 615-20.

Fare, R., Grosskopf, S., and Hernandez-Sancho, F. (2004a) "Environmental performance: an index number approach", Resource and Energy Economics, Vol 26 No 4, 343-52.

Fare, R., Grosskopf, S., and Weber, W. L. (2004b) "The effect of risk-based capital requirements on profit efficiency in banking", Applied Economics, Vol 36 No 15, 1731-43.

Hansen, J. W. and Jones, J. W. (1996) "A systems framework for characterizing farm sustainability", Agricultural Systems, Vol 51 No 2, 185-201.

Herring, H. and Roy, R. (2002) “Sustainable services, electronic education and the rebound effect”, Environmental Impact Assessment Review 22(5), 525-542.

IPCC (2007). Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Kuosmanen, T., et al. (2009). “Firm and industry level profit efficiency analysis using absolute and uniform shadow prices”. European Journal of Operational Research, doi:10.1016/j.ejor.2009.06.002

KWIN (2007). Kwantitatieve Informatie Veehouderij 2007-2008. Handboek 2. Animal Science Group, Wageningen UR, pp408. ISSN 1570-8594.

Lawn, P. (2007). Frontier issues in ecological economics, Edward Elgar, Cheltenham, UK, 374 pp. ISBN 978 1 84542 840 2

Lee, J. D., Park, J. B., and Kim, T. Y. (2002) "Estimation of the shadow prices of pollutants with production/environment inefficiency taken into account: a nonparametric directional distance function approach", Journal of Environmental Management, Vol 64 No 4, 365-75.

Li, S. and Ng, Y.C. (1995). “Measuring the productive efficiency of a group of firms”. International Advances in Economic Research, Vol 1, No 4, 377-390.

Lozano, S., Villa, G., and Brannlund, R. (2009) "Centralised reallocation of emission permits using DEA", European Journal of Operational Research, Vol 193 No 3, 752-60.

Luenberger, D. G. (1992) "Benefit functions and duality", Journal of Mathematical Economics, Vol 21 No 5, 461-81.

Mayumi, K., Giampietro, M. and Gowdy, J. M. (1998), `Georgescu-Roegen/Daly versus Solow/Stiglitz revisited', Ecological Economics 27(2), 115-117.

McMullen, B. S. and Noh, D. W. (2007) "Accounting for emissions in the measurement of transit agency efficiency: A directional distance function approach", Transportation Research Part D-Transport and Environment, Vol 12 No 1, 1-9.

Murty, M. N., Kumar, S., and Dhavala, K. K. (2007) "Measuring environmental efficiency of industry: a case study of thermal power generation in India", Environmental & Resource Economics, Vol 38 No 1, 31-50.

Reinhard, S., Lovell, C. A. K., and Thijssen, G. (1999) "Econometric estimation of technical and environmental efficiency: An application to Dutch dairy farms", American Journal of Agricultural Economics, Vol 81 No 1, 44-60.

Seidl, I. and Tisdell, C.A. (1999) "Carrying capacity reconsidered: from Malthus' population theory to cultural carrying capacity", Ecological Economics, Vol 31 No 3, 395-408.

Tyteca, D. (1996) "On the measurement of the environmental performance of firms - A literature review and a productive efficiency perspective", Journal of Environmental Management, Vol 46 No 3, 281-308.

Voinov, A. and Farley, J. (2007) "Reconciling sustainability, systems theory and discounting", Ecological Economics, Vol 63 No 1, 104-13.