Economic efficiency when prices are not fixed: Disentangling quantity and price efficiency

Maria Conceição A. Silva Portela [*1], Emmanuel Thanassoulis [**]

* CEGE - Centro de Estudos em Gestão e Economia, Faculdade de Economia e de Gestão, Centro Regional do Porto da Universidade Católica Portuguesa

** Aston University, B4 7ET Birmingham, UK

Abstract

This paper proposes an approach to compute and decompose cost efficiency in contexts where units can adjust input quantities and seek input prices so that through the joint determination of quantities and prices they can minimise the aggregate cost of the outputs they secure. The models developed are based on the data envelopment analysis framework and can accommodate situations where the degree of influence over prices is minimal, and situations where there is a large influence over prices. When units cannot influence input prices the models proposed reduce to the standard cost efficiency, where prices are taken as exogenous. In addition to the cost model, we introduce a novel decomposition of cost efficiency into a quantity and a price component, based on Bennet indicators. The components are expressed in terms of percentage cost savings that can be attained through changing prices and changing quantities towards the overall optimum cost target.

Keywords: Cost efficiency; price efficiency, Bennet Indicators, Data Envelopment Analysis

1 Introduction

Traditional models for computing cost and revenue efficiency date back to Farrell (1957) and will be called Farrell cost or revenue efficiency models. Since the appearance of Data Envelopment Analysis (DEA) in 1978 (see Charnes et al. (1978)) cost efficiency and revenue efficiency have been computed through linear programming

---

[1] *Correspondence: Maria Conceição A. Silva Portela , Faculdade de Economia e Gestão da Universidade Católica Portuguesa, Rua Diogo Botelho, 4169-005 Porto, Portugal.* E-mail: csilva@porto.ucp.pt, Tel. +351226196200

models, when an option for the use of non-parametric models is taken. The alternative is to compute cost or revenue efficiency based on the definition of parametric revenue functions or cost functions like the Cobb Douglas function or the translog function (see e.g. Greene (2008)). In both cases, the underlying economic model of the firm typically assumes that it operates in a competitive market, where prices of inputs and outputs are exogenously given, often taken at the level of actual prices observed at the operating decision making unit (DMU). As a result, cost efficiency and revenue efficiency reflect cost savings or revenue gains that can accrue from changes in input or output quantities given their (fixed) prices. Units where input or output prices are exogenously fixed are said to be *price takers*.

In this paper we depart from the notion of production units being price takers and consider the case where the units to be assessed for cost or revenue efficiency either can influence to some extent the input or output prices they secure and/or could secure prices other than those they actually did even if they do not influence their levels directly themselves. This in turn means that such production units can improve their cost or revenue efficiency by inter alia securing better prices for their inputs and/or outputs. We focus on cost efficiency in this paper, but the extension of the proposed approach to revenue efficiency is straightforward.

Modelling economic efficiency in situations where units are not price takers has been addressed before, **but mainly in a context of endogenous prices (i.e. the level of outputs of a DMU are assumed to impact the price per unit of output, so that maximum output levels may not necessarily be compatible with maximum revenue)**. In such contexts prices are said to be endogenous to the unit. For example, Cherchye et al. (2002) considered a situation of endogenous and uncertain prices, while Johnson and Ruggiero (2011) considered the situation of endogenous prices. On the other hand, Kuosmanen and Post (2002) considered just the situation of price uncertainty. Price endogeneity as modelled hitherto has relied on some explicit functional form e.g. between output quantities and corresponding output prices. We depart in this paper from this modelling paradigm for two reasons. Firstly, to deal with situations where markets are not competitive and demand functions are not known, and secondly even if the markets are perfectly competitive the units concerned may not be able to access the fully competitive prices which in any case may not necessarily be explicitly known.

We would argue that prices manifested at unit level apart from endogeneity or exogeneity as the case may be, also incorporate an element of managerial efficiency and local factors, which can make the task of predicting prices attainable at unit level very difficult. That is, whether prices are endogenous or exogenous the ability of a unit to secure optimal prices reflects, to an extent, also the unit's ability to assess the pricing context in which it is operating and take appropriate action to optimise the prices it secures. In this context we take the prices manifested across the DMUs as the best available evidence of prices that can be secured, and we replicate in the context of input prices what is done for input-output levels in a non parametric context of efficiency measurement. In traditional DEA models no assumption is made of the functional relationship between input and output levels. The production possibility set is built with reference to observed input-output level correspondences using certain assumptions (eg. Thanassoulis et al. (2008, p.255)). In the same manner in this paper we use observed input prices to derive input prices that are in principle attainable by DMUs. The larger the number of DMUs within a 'pricing environment' the better the prices attainable by DMUs will be revealed.

To see how manifested input prices can incorporate some component of inefficiency, even in competitive markets, consider the case of a number of hypothetical hospitals in a given large city each one looking to hire a doctor of the same skills at the same point in time. It is by no means certain what the theoretically minimum salary at which such a doctor can be hired is. The salary each hospital will end up paying to each doctor recruited will to some extent depend on the negotiating skills of the candidate and of the employer, on whether the post was over or under specified relative to the skills actually needed and on externalities such as location of the post relative to a candidate's residence, perceived culture of the employer etc. Similarly, in a banking context, where input prices can be interest paid on deposits, the ability of management to devise financial products in terms of the interplay of interest rate and withdrawal or other restrictions will affect the rate at which a bank can secure funds, whether or not there is a free competitive banking market. Thus different banks drawing funds from the same pool of savings can secure funds at different rates in effect meaning securing inputs at different prices. In both these examples the actual salaries for doctors paid or the interest rates secured by banks offer us the best available empirical evidence of the prices that might have been

attainable. As in the case of input levels that can secure given output levels, in the case of prices too, the more the units in the comparative set the more likely we are to observe the most favourable prices that could have been secured in a given operating context.

Thus in this paper we address situations where units are not strictly input price takers and prices secured reflect inter alia managerial effectiveness in securing optimal prices for their inputs. We argue that when the performance of such units is assessed, managerial ability to arrive simultaneously at an optimal price and quantity mix so as to minimise costs should be captured. We develop a cost model for the simultaneous optimisation of input quantities and prices. Such a model requires that data on prices and quantities are available, and assumes that DMUs have some degree of influence both over prices and over quantities of inputs.

The paper is structured as follows. In the next section we review literature that is more directly related with the work developed in this paper, and will further address the motivations of the paper. In section 3 we propose a new model for computing cost efficiency and show how to decompose this measure in section 4. In section 5 an illustrative example is shown highlighting the differences between ours and existing approaches for computing price efficiency. Section 6 concludes the paper.

2   Economic efficiency in non-competitive markets

Consider for each DMU j ($j = 1, ..., n$) a vector $\mathbf{x_j} = (\mathbf{x_{1j}, x_{2j}, ..., x_{mj}})$ reflecting m inputs consumed for producing a vector of s outputs $\mathbf{y_j} = (\mathbf{y_{1j}, y_{2j}, ...y_{sj}})$. Consider also, that observed prices of inputs at DMU_j are known and given by a vector $\mathbf{p_j} = (\mathbf{p_{1j}, p_{2j}, ..., p_{mj}})$. Observed aggregate cost of inputs for a given unit o is $C_o = \sum_{i=1}^{m} p_{io}x_{io}$. The Farrell cost efficiency model for DMUo is the solution of the linear program in (1), where input quantities, x_i, and the intensity variables, λ_j, are taken as the decision variables and prices are considered exogenous (see e.g. Färe et al. (1985)).

$$\min_{\lambda_j, x_i} \Big\{ C = \sum_{i=1}^{m} p_{io}\, x_i \mid \sum_{j=1}^{n} \lambda_j x_{ij} \leq x_i,\ i = 1, ..., m\,,\ \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro},\ r = 1, ..., s, \lambda_j, x_i \geq 0 \Big\}$$

$$(1)$$

The optimal solution to model (1) yields the minimum cost (C^*), at which DMU_o can secure at least its output levels when input prices are taken as given. The model in (1), also yields the optimal input quantities (x_i^*) which support the outputs of DMU_o at the minimum cost (C^*). Cost efficiency is defined as the ratio of the minimum cost to the observed cost (C^*/C_o).

Traditionally reduction in costs is prescribed in two ways taken in sequence: (i) reducing the quantities of inputs pro-rata to reach the minimum input levels capable of supporting the outputs; (ii) changing the mix of inputs so that, given the prevailing input prices, aggregate input costs are minimised. The technical efficiency (computed using the standard DEA model of Charnes et al. (1978)) of a unit reflects its scope for savings through (i) and its allocative efficiency is reflected through (ii).

Usually minimum cost models are computed under constant returns to scale (CRS), as in (1), but both the cost and the technical efficiency measure can also be computed under variable returns to scale VRS (see e.g. Färe et al. (1994) and Athanassopoulos and Gounaris (2001)).

Tone (2002) noted that the Farrell cost model does not capture the full extent of cost savings, as the cost efficiency (C^*/C_o) of two DMUs may be equal, even if one faces double the input prices of the other (as long as both show the same levels of inputs and outputs). In essence, if prices are truly exogenous this outcome is acceptable, since both DMUs use the same mix of inputs for the prevailing input prices, notwithstanding the fact that one of the DMUs has to pay twice as much for its inputs as the other. In a reply to Tone (2002), Färe and Grosskopf (2006) propose the use of a difference form of cost efficiency, rather than a ratio form, to solve the problem. Such a difference form is based on the notion of directional distance functions and implies the use of a common normalisation factor. In spite of apparently solving the problem, the approach of Färe and Grosskopf (2006), provides for two DMUs employing the same quantities of inputs to produce the same quantities of outputs the same technical inefficiency, but an allocative inefficiency that is for one DMU double that of the other, when input prices are also double. However, this allocative inefficiency does not reflect what is traditionally reflected by an allocative efficiency component (i.e. the extent to which the mix of inputs needs to be changed given the prevailing input prices) since the mix changes in

inputs are exactly the same for both DMUs. Such allocative inefficiency is in fact reflecting price differences, which the units are assumed not to control, and therefore this component should be ignored on the cost inefficiency measurement.

Clearly, if one assumes, as we argue, that production units have some degree of influence over input prices either because of an uncompetitive market and/or because they can fail to secure optimal prices that may have been available, then this should be taken into account when measuring cost efficiency. Indeed it could be argued that even if units have no control over prices they may still be interested in knowing the extent to which they fail to reach minimum cost due to suboptimal prices as distinct from technical inefficiency. **In this case, such component of inefficiency should not be called allocative inefficiency (as in Färe and Grosskopf (2006)), but should be called price inefficiency**. Tone and Tsutsui (2007), following Tone (2002), proposed a decomposition of cost efficiency into technical, allocative and price efficiency. The price efficiency component of Tone and Tsutsui (2007) reflects the scope for savings through input price changes, whereas allocative efficiency is defined as "the adjustment to the optimal cost mix, viz., the combination of the optimal input amount and input price mixture" (Tone and Tsutsui (2007) , p. 95). These two concepts in the Tone and Tsutsui (2007) model are mis-specified, as allocative efficiency is unrelated to its traditional meaning and price efficiency does not capture entirely the changes in prices as we will see later. (see also Sahoo and Tone (2013) for a recent application).

Camanho and Dyson (2008) also addressed the situation of non-competitive markets where prices can be negotiated rather than imposed by a theoretical market equilibrium. Their approach, starts by the computation of the traditional Farrell cost efficiency measure and its decomposition into technical and allocative components, and then identifies a third component called 'market efficiency'. This component is identified by solving the traditional cost efficiency model under different price assumptions (one of which is the assumption that the minimum price observed on each input can be attained by all DMUs, and another is that each DMU can choose from the set of observed price vectors the one that minimises the aggregate cost of inputs for their output bundle). Ray et al. (2008) also proposed a related method for modelling situations where firms can choose their location depending on the input prices offered in each location. The main novelty in the Ray et al. (2008)

model is that it considers the possibility of partially producing an output bundle from an input bundle in different locations and at different prices.

A related strand of the literature is that dealing with price uncertainty or incomplete price information. Such literature generally makes use of the dual of technical efficiency measures, since it is well known in economics that the Farrell technical efficiency can be interpreted as a cost minimisation model under the most favourable shadow prices. To see this, consider the dual of model (1), shown in (2), where v_i are the input shadow prices and u_r are the output shadow prices.

$$\max_{u_r, v_i} \left\{ \sum_{r=1}^{s} u_r \, y_{ro} \mid -\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} \leq 0, \; j = 1, ..., n \,, v_i \leq p_{io}, v_i, u_r \geq 0 \right\} \quad (2)$$

At the optimal solution of model (1) the variables x_i are basic variables (as $x_i \geq \sum_{j=1}^{n} \lambda_j x_{ij}$) and therefore the corresponding constraints $(v_i \leq p_{io})$ in the dual (2) are binding, meaning that all input weights or shadow prices are equal to observed input prices $(v_i = p_{io})$. **If observed prices p_{io} cannot be known, but some information on prices is available (e.g. the price of input 1 is more than the double of that of input 2) model (1) cannot be used, but its dual (2) can be used by replacing constraints $(v_i \leq p_{io})$ by other constraints reflecting the price information available for inputs (e.g. $v_1 \geq 2v_2$). Such additional constraints will led to a solution that is an approximation to the economic efficiency measure in (1).** The dual of the Debreu-Farrell model of technical efficiency is similar to (2), but the shadow price constraints $(v_i \leq p_{io})$ are replaced by a price normalisation constraint $\sum_{i=1}^{m} v_i x_{io} = 1$. This constraint is obviously less strict and allows a wide range of choice in optimal shadow prices, rather than equating them to observed prices. Therefore, the technical efficiency measure can be seen as an upper bound of cost efficiency (see e.g. Russell (1985) or Leleu (2013)). Several authors have used this approach to model economic efficiency, such as Camanho and Dyson (2005), Cherchye et al. (2002) or Kuosmanen et al. (2010), where price uncertainty has been modelled through constraints imposed on the shadow prices in model (2). (2) is specified in 'price space' and shadow prices are the decision variables to this model. Since we also use prices as decision variables, this strand of the literature is related to ours. However, optimisation of shadow prices differs from optimisation of observed prices, which is what we pursue in this paper

(note that we assume that observed prices can be changed or optimised, and therefore are not exogeneous). In fact shadow price optimization, should be attempted when price information is incomplete or unavailable. When price information is available, one can optimise over observed prices, if one assumes that DMUs have some degree of influence over the input prices they face. Therefore based on two assumptions: that data on observed input prices are available and that DMUs can to some extent influence them, we depart from existing cost efficiency models in two fundamental ways. Firstly we allow quantities of inputs and input prices to vary simultaneously, and secondly we decompose cost efficiency so as to capture separately the savings due to exploiting input quantity changes and those due to exploiting input price flexibility. We argue this decomposition is more appropriate when there is input price flexibility rather than the notion of allocative efficiency used for example in Camanho and Dyson (2008). In fact traditional allocative efficiency is not very meaningful when one assumes that there is an optimum level for prices, and that units can make efforts to change prices accordingly to this optimum level.

3 A cost efficiency model when input quantities and their prices can vary simultaneously

The new cost minimisation model proposed in this paper is in all respects similar to the Farrell cost model in (1) except that it assumes each DMU is not a price taker, but rather it could have in principle availed itself of input prices other than those it actually paid. Under this scenario a modification of model (1) is needed to the form depicted in (3).

In (3) decision variables reflect changes in observed levels of inputs (θ_i), changes in observed levels of prices (γ_i), and the intensity variables associated to DMUj, as in traditional DEA models. We also introduce the decision variables z_{ij}, the ijth being associated with price i observed at DMUj. This set of additional intensity variables, as will be explained shortly, makes it possible to arrive at a set of input prices that might have been available to the DMU being assessed, by using the data on input prices that have been observed across the set of DMUs. Thus the new set of intensity variables z_{ij} play a similar role to that of the intensity variables λ_j albeit in respect of input prices rather than input-output levels.

$$\min_{\gamma_i, \theta_i, \lambda_j, z_{ij}} \left\{ C = \sum_{i=1}^{m} \gamma_i p_{io} \, \theta_i x_{io} \mid \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_i x_{io}, \, i = 1, ..., m, \; \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, r = 1, ..., s, \right.$$

$$\left. \sum_{j=1}^{n} z_{ij} \, p_{ij} \leq \gamma_i p_{io}, i = 1, ..., m \; \sum_{j=1}^{n} z_{ij} = 1, \; i = 1, ..., m, \; z_{ij}, \lambda_j, \theta_i, \gamma_i \geq 0 \right\} \quad (3)$$

The model in (3) is non-linear in the objective function in that the sum of the product of input quantities and prices is minimised (note that $\theta_i x_{io}$ can be replaced by a single variable x_i and that $\gamma_i p_{io}$ could be replaced by a variable p_i). In this manner input prices and quantities are optimised simultaneously. Model (3) has two types of constraints: those defining feasible input-output correspondences and those defining feasible input prices.

The feasible input-output correspondences in model (3) are the same as those in model (1), defining the production possibility set in the traditional DEA model under CRS, (though this can be readily modified for VRS technologies by adding the convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$).

The constraints on prices $\sum_{j=1}^{n} z_{ij} \, p_{ij} \leq \gamma_i p_{io}, i = 1, ..., m$ and $\sum_{j=1}^{n} z_{ij} = 1, \; i = 1, ..., m$ state that the feasible in principle price for each input would be a convex combination of observed prices for that input $(p_i \geq \sum_{j=1}^{n} z_{ij} \, p_{ij})$. The rationale behind the price constraints is that the best guide we have of input prices available to DMUs are those that have been observed.

The optimal solution from model (3) renders input quantity targets $(x_i^* = \theta_i^* x_{io})$, price targets $(p_i^* = \gamma_i^* p_{io})$, benchmarks for input-output quantities (all units j whose $\lambda_j > 0$), and benchmarks for each input price i, (all units j whose $z_{ij} > 0$). Note that in this general formulation, we assume that referent DMUs for optimal input prices do not need to be the same for all inputs nor do they need to be the same as the benchmarks for technically efficient input-output levels. That is, there is no reason to assume that a unit with scope for savings would need to emulate the same peer units for input prices as for transforming inputs to outputs. Further, even if the same benchmarks are found for both input to output transformations and for input price emulation, the intensities with which they form the virtual comparator unit need not be the same in both cases. In standard DEA a virtual unit is formed using the same intensities (λ_j) for inputs and outputs, as there is a causal correspondence between input and output quantities. Such causal correspondence is not assumed in

model (3) between input prices and input or output quantities. However, in cases of endogeneity (e.g. where input price advantages may be secured through volume purchases) additional constraints can be placed in model (3) linking the volume of input of the virtual comparator unit with the input prices feasible in principle. Indeed other prior information about constraints on available prices for one or more inputs can be included in model (3), if available. However, again for reasons of simplicity we avoid at this stage refinements of this type to focus on the departure from the traditional price taker modelling approach prevailing so far in the literature.

The optimal solution to model (3) is in fact a trivial one, as far as prices are concerned. Optimum prices in (3) are the minimum observed prices for each input. Thus in a sense the model in (3) is equivalent to the approach in Camanho and Dyson (2008) where they directly opt to use the minimum observed price for each input as the basis for computing the cost efficiency of each unit in the framework of the traditional cost model (1). However, this is only so for the least restricted case on input prices as depicted in model (3). Model (3) is to be seen as a generic one, where the Farrell cost efficiency model and the approaches developed by Camanho and Dyson (2008) are special cases. Our formulation in (3) can yield the same results as the Farrell cost efficiency model (1) when we set $\gamma_i = 1$ for all i. In this case the constraints on prices would become $\sum_{j=1}^{n} z_{ij} p_{ij} \leq p_{io}$, and this would be redundant (as for the unit under assessment z_{io} can be set to 1 without violating any constraints). On the other hand, if we assume input prices are related across inputs so that only convex combination of the vectors of observed prices rather than of individual input prices are feasible, then model (3) can be modified by using intensity variables z_j, instead of z_{ij}. This modification will identify one of the observed vectors of input prices as optimal as was the case in one of the approaches recommended by Camanho and Dyson (2008).

Thus the model in (3) can be seen as the most general model in which optimal input quantities and prices are to be determined ranging from the case where input prices are fixed and only optimal quantities are to be determined, to the case where there are in effect no restrictions on input prices other than they should be higher than some convex combination of observed prices, in which case the lowest observed price on each input is optimal. Restrictions on individual prices or combinations of prices can be readily accommodated in (3) depending on context. We illustrate

here one simple restriction which we would expect to be valid in many real life cases. It would be reasonable to assume that a DMU has on going relationships with suppliers. Thus, the convex combination of observed prices on an input within model (3) should be further restricted so that DMUo could not be expected to secure input prices too far removed from those it has secured in the past, at least not in the short term. The scope for potential reductions in input prices would be decision maker supplied or drawn from the experience of reductions in input prices other DMUs may have secured. Model (3) can be readily modified to cope with this type of restriction by adding the constraint $\gamma_i \geq \alpha_i$ where α_i is a user-defined scalar between 0 and 1 providing a lower bound for the level of the price of input i at DMU_o.

We conclude this section by noting that the above model is non-linear, requiring the use of advanced non linear programming solvers for obtaining a solution that is a global optimum. Developments in the field of non-linear programming have now reached a certain level of maturity (see Pintér (2007)) which permits the solution of models such as (3) to identify a global optimal solution. Traditional solvers that could only guarantee local optimal solutions to non-linear models are being replaced by more efficient solvers which perform global scope searches and can reach global or very close to global optimal solutions. In Pintér (2007) the authors explain the use of a Gams solver LGO to solve non-linear models and illustrate the performance of this solver. Given the large availability of solver options in e.g. Gams indeed trying more than one solver is the best option to guarantee obtaining, **or being close** to the global optimal solution.

4 Decomposing the cost efficiency measure when units are not input price takers

Through model (3) a cost efficiency measure can be computed as the ratio between minimum cost and observed cost. Such a ratio is traditionally decomposed into technical and allocative efficiency measures, but such a decomposition lacks meaning when both prices and quantities are decision variables. As noted earlier, the decompositions proposed in previous attempts to model non-competitive markets (such as the Camanho and Dyson (2008) or the Tone and Tsutsui (2007) approaches),

are in our view ill-defined in a context where both prices and quantities can vary simultaneously to minimise costs. We propose instead to identify the components of cost savings attributable to input price changes and those attributable to input quantity changes making use of price and quantity indicators as defined in Balk, et al. (2004) and Diewert (2005). Such indicators are usually defined in temporal analyses, where cost or revenue change over time is decomposed into volume and price change (see Diewert (2005)). We follow a similar procedure in our decomposition of cost efficiency, but comparing observed versus optimal quantities and prices.

Model (3) yields the minimum cost value (C^*) and the corresponding optimal quantity and price targets. The ratio between minimum cost and observed cost is our measure of cost efficiency (CE):

$$CE = \frac{C^*}{C_o} = \frac{\sum_{i=1}^{m} \theta_i^* \gamma_i^* x_{io} p_{io}}{\sum_{i=1}^{m} x_{io} p_{io}} \qquad (4)$$

This ratio cannot be readily disentangled into quantity and price contributions to aggregate potential savings because these vary by input. We can compute, however, a cost inefficiency measure (1- CE) reflecting the percentage of observed aggregate input costs which can be saved through adopting the target input prices ($p_{io}^* = \gamma_i^* p_{io}$) and quantities ($x_{io}^* = \theta_i^* x_{io}$) for each unit o. This can be expressed as:

$$\frac{\sum_{i=1}^{m} x_{io} p_{io} - \sum_{i=1}^{m} x_{io}^* p_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} = 1 - \frac{\sum_{i=1}^{m} x_{io}^* p_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} = 1 - CE \qquad (5)$$

This can be decomposed into:

$$\frac{\sum_{i=1}^{m} x_{io} p_{io} - \sum_{i=1}^{m} x_{io}^* p_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} = \frac{\sum_{i=1}^{m} (x_{io} - x_{io}^*) p_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} + \frac{\sum_{i=1}^{m} (p_{io} - p_{io}^*) x_{io}}{\sum_{i=1}^{m} x_{io} p_{io}} \qquad (6)$$

In (6) changes in quantities are weighted by optimal prices and changes in prices are weighted by observed quantities. Clearly we could choose another set of weights as shown below:

$$\frac{\sum_{i=1}^{m} x_{io} p_{io} - \sum_{i=1}^{m} x_{io}^* p_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} = \frac{\sum_{i=1}^{m} (x_{io} - x_{io}^*) p_{io}}{\sum_{i=1}^{m} x_{io} p_{io}} + \frac{\sum_{i=1}^{m} (p_{io} - p_{io}^*) x_{io}^*}{\sum_{i=1}^{m} x_{io} p_{io}} \qquad (7)$$

Since the choice of weights is arbitrary we can sum both expressions and arrive at the following equality:

$$\frac{\sum_{i=1}^{m} x_{io}p_{io} - \sum_{i=1}^{m} x_i^* p_{io}^*}{\sum_{i=1}^{m} x_{io}p_{io}} = \frac{\sum_{i=1}^{m} (x_{io} - x_{io}^*)(\frac{p_{io}+p_{io}^*}{2})}{\sum_{i=1}^{m} x_{io}p_{io}} + \frac{\sum_{i=1}^{m} (p_{io} - p_{io}^*)(\frac{x_{io}+x_i^*}{2})}{\sum_{i=1}^{m} x_{io}p_{io}} \quad (8)$$

The first component in the RHS of (8) represents the percentage of observed costs that can be saved through changing input quantities. Similarly the second component in (8) represents the percentage of observed costs that can be saved through changing input prices. The decomposition presented above can be seen in the literature in other contexts. In particular the decomposition in (8) is a Bennet indicator where **quantity differences are evaluated at average prices and price differences are evaluated at average quantities** (see e.g. Chambers (2002) or Grifel-Tatjé and Lovell (2000, 1999) who used similar decompositions but in a context of cost and profit change over time decomposed into a quantity effect and a price effect). Though known, these indicators have never been used, to the authors' knowledge, in the context of decomposing economic measures of efficiency as we do in this paper. The axiomatic theory on Bennet's indicators was explored by Diewert (2005) who showed that the Bennet price and quantity indicators satisfy all of the 18 tests defined in his paper (continuity, identity, monotonicity in prices and quantities, units invariance, linear homogeneity, to name but a few.)

Our approach, therefore requires solving just one cost minimisation problem for each firm, where prices and quantities are assumed as decision variables, and then, makes use of Bennet indicators to provide a decomposition of total cost savings into those that can result from changes in prices and those that can result from changes in quantities. Note that each one of the components in (8) can be further decomposed into a radial and mix component in a manner which can make contact with the traditional notions of radial technical and allocative efficiency in classical cost decompositions. To proceed with this decomposition we first use the input price and quantity targets obtained directly from (3) to compute radial components (x_{io}^{*R} and p_{io}^{*R}) representing respectively the feasible pro-rata change of input quantities and prices. The remainder of the feasible changes in input quantities and prices needed to attain the minimum aggregate cost are non-radial, reflecting mix changes in input quantities and prices. The decomposition then obtained for input quantities into a radial and a mix component of cost savings is shown in (9), and that obtained for input prices is shown in (10).

$$\frac{\sum_{i=1}^{m}(x_{io}-x_{io}^{*})(\frac{p_{io}+p_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}} = \frac{\sum_{i=1}^{m}(x_{io}-x_{io}^{*R})(\frac{p_{io}+p_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}} + \frac{\sum_{i=1}^{m}(x_{io}^{*R}-x_{io}^{*})(\frac{p_{io}+p_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}}$$

$$(9)$$

$$\frac{\sum_{i=1}^{m}(p_{io}-p_{io}^{*})(\frac{x_{io}+x_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}} = \frac{\sum_{i=1}^{m}(p_{io}-p_{io}^{*R})(\frac{x_{io}+x_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}} + \frac{\sum_{i=1}^{m}(p_{io}^{*R}-p_{io}^{*})(\frac{x_{io}+x_{io}^{*}}{2})}{\sum_{i=1}^{m}x_{io}p_{io}}$$

$$(10)$$

Where,

$$x_{io}^{*R} = \begin{cases} x_{io}(\max_i \theta_i^*) & \text{if } \theta_i^* \leq 1 \quad \forall i \\ x_{io}(\min_i \theta_i^*), & \text{if } \theta_i^* \geq 1 \quad \forall i \\ x_{io} & \text{all other cases} \end{cases} \quad (11)$$

and

Where

$$p_{io}^{*R} = \begin{cases} p_{io}(\max_i \gamma_i^*) & \text{if } \gamma_i^* \leq 1 \quad \forall i \\ p_{io}(\min_i \gamma_i^*), & \text{if } \gamma_i^* \geq 1 \quad \forall i \\ p_{io} & \text{all other cases} \end{cases} \quad (12)$$

That is, when all optimal θ_i^* and γ_i^* have values below 1 we obtain radial components (x_{io}^{*R} and p_{io}^{*R}) directly from the solution of model (3) by multiplying each observed input quantity or price by the maximum of the optimal θ_i^* and γ_i^*, respectively. In the unlikely event (since costs are being minimised) that all optimal θ_i^* or γ_i^* are above 1, then the radial components will be obtained using the lowest θ_i^* and γ_i^* values. Finally, where optimal θ_i^* and γ_i^* values span the range below and above 1 we have a mix of contraction and expansion of input quantities and prices which means we have no traditional radial component and all savings are due only to mix of input quantity and price changes.

5 Illustrative example

In order to illustrate our approach and contrast it with existing approaches for computing and decomposing cost efficiency when units are not input price takers, we use the illustrative example in Table 1, also depicted in Figure 1.

Table 1: Ilustrative example

Unit	X_1	X_2	Y	P_1	P_2	C_1	C_2	$C_o = C_1 + C_2$
A	3	2	1	2	3	6	6	12
B	3	2	1	4	0.5	12	1	13
C	2	1	1	2	4	4	4	8
D	6	4	1	1	2	6	8	14
E	1	2	1	2	3	2	6	8

Units E and C are technically efficient. Unit D has double the levels of inputs of unit A to produce the same output and therefore is half as technically efficient as A. Units A and B use the same input quantities, but buy them at different prices, whereas units A and E use different input quantities but buy them at the same prices. We therefore expect that units A and E are equally efficient in terms of prices, when quantities are ignored, and that units A and B are equally efficient in terms of quantities, when prices are ignored. In addition, unit C faces double input prices compared to unit D and therefore should have half the price efficiency of D, when quantities are ignored. As we shall see later, having the same prices or the same quantities should result in the same optimal changes in quantities or prices, but not necessarily in the same efficiency measures, when efficiency is expressed in cost saving terms.

We argue in this paper that our proposed approach has mainly two advantages: **(i) it offers a clear decomposition of an overall scope for cost savings into quantity and price gains, while existing approaches use concepts of allocative efficiency that lack meaning in a context of non-exogenous prices; and (ii) it can include explicit price relationships, in a mathematical model, providing price targets that are more realistic than those in existing models.** In what follows we will use the above example to illustrate both these points.

5.1 A clear decomposition of overall cost savings

Figure 1 depicts the traditional representation of DMUs in the input-output space, where DMUs E and C form the technical efficient frontier (they show the minimum input quantities given outputs produced) and units A, B and D are technically inefficient. We also represent in this figure the isocost line of unit D, under its observed input prices. The minimum cost for unit D, under its observed prices
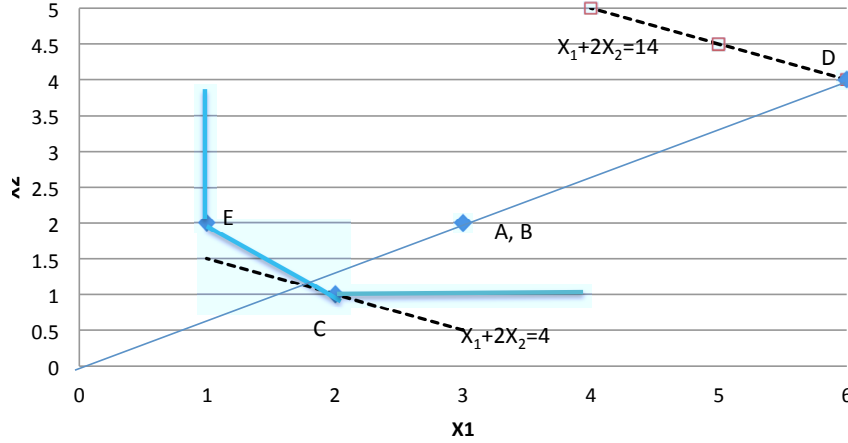
Figure 1: Isoquant and isocost for illustrative example.

is 4 and occurs at the input quantities and mix of unit C. Therefore the Farrell cost efficiency of unit D (i.e. under its observed input prices) is given by $4/14 = 28.57\%$ and this measure can be decomposed into a technical efficiency measure (30%) and an allocative efficiency measure (95.24%).

In Table 2 we show the results for the Farrell cost efficiency model (FM), the Tone and Tsutsui (2007) model (TT), the Camanho and Dyson (2008) model(CD), and our (PT) model, as resulting from (3) and decomposition in (8). Note that for Camanho and Dyson (2008) we distinguish two approaches (i), where it is assumed that the minimum price observed for each input across the units is actually available to all units and so units are assessed in relation to this price vector which normally would not have been observed at a single unit, and (ii) where it is assumed that only observed combinations of prices are possible and each unit is assessed in relation to all observed combinations of prices and the set giving it the lowest aggregate input cost is chosen. Note also that efficiency values shown in Table 2 are not directly comparable, since existing approaches decompose efficiency scores whereas the PT approach developed in this paper decomposes cost savings (and therefore inefficiency). Note however that efficiency scores can be read as potential for cost savings since an efficiency score of 80% means that costs can be reduced to 80%

of observed levels, meaning that costs can be reduced by 20%. We decided not to convert values to a common basis because the multiplicative or additive nature of decompositions would not be easily seen if we have done such a conversion.

Table 2: Results from various approaches

		A	B	C	D	E
FM	Observed Cost	12	13	8	14	8
	Min Cost	7	5	8	4	7
	Targets (x1, x2)	(2, 1)	(1, 2)	(2, 1)	(2, 1)	(2, 1)
	Technical eff	60%	60%	100%	30%	100%
	Allocative eff	97.22%	64.10%	100%	95.24%	87.50%
	Farrell Cost Eff	58.33%	38.46%	100%	28.57%	87.50%
TT	Min Cost	4.2	4.2	4.2	4.2	4.2
	Targets (C1, C2)	(1.8, 2.4)	(1.8, 2.4)	(1.8, 2.4)	(1.8, 2.4)	(1.8, 2.4)
	Technical eff	60%	60%	100%	30%	100%
	Price eff	62.50%	100%	56.25%	100%	52.50%
	Allocative eff	93.33%	53.85%	93.33%	100%	100%
	Cost Eff	35%	32.31%	52.50%	30%	52.50%
CD (i)	Min Cost	2	2	2	2	2
	Targets (x1, x2)*	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
	Targets (p1, p2)	(1, 0.5)	(1, 0.5)	(1, 0.5)	(1, 0.5)	(1, 0.5)
	Farrell cost eff	58.33%	38.46%	100%	28.57%	87.5%
	Price eff**	28.57%	40%	25%	50%	28.57%
	Cost Eff	16.67%	15.38%	25%	14.29%	25%
CD(ii)	Min Cost	4	4	4	4	4
	Targets (x1, x2)*	(2, 1)	(2, 1)	(2, 1)	(2, 1)	(2, 1)
	Targets (p1, p2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
	Farrell cost eff	58.33%	38.46%	100%	28.57%	87.5%
	Price eff**	57.14%	80%	50%	100%	57.14%
	Cost Eff	33.33%	30.77%	50%	28.57%	50%
PT	Min cost	2	2	2	2	2
	Targets (x1, x2)	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)
	θ_1	0.333	0.333	0.5	0.167	1
	θ_2	1	1	2	0.5	1
	Targets (p1, p2)	(1, 0.5)	(1, 0.5)	(1, 0.5)	(1, 0.5)	(1, 0.5)
	γ_1	0.5	0.25	0.5	1	0.5
	γ_2	0.1667	1	0.125	0.25	0.1667
	Potential savings due to quantity changes	25.00%	38.46%	-9.38%	53.57%	0.00%
	Potential savings due to price changes	58.33%	46.15%	84.38%	32.14%	75.00%
	Total potential savings	83.33%	84.62%	75.00%	85.71%	75.00%
	Cost efficiency	16.67%	15.38%	25.00%	14.29%	25.00%

* quantity targets in the CD approach are for the second stage models as the first stage targets are those from the traditional cost efficiency approach (FM)

** We replaced Market efficiency by price efficiency in the CD approach for ease of comparison

Results in Table 2 show that under the Farrell cost efficiency model (FM) unit C is overall cost efficient, and the remaining units have unit C as their benchmark

for cost minimisation, except unit B (since the price of input 2 for unit B is 8 times lower than the price of input 1, the cost minimising point for unit B is point E and not point C). This approach assumes that prices are exogenous, and as a result the fact that unit's C prices are twice those of unit D is not reflected in the efficiency score of this unit.

Approaches TT, CD and PT assume that prices can be to some extent influenced through managerial actions. Under these approaches units C and E are the most cost efficient units (they are shown equally efficient under TT, CD and PT approaches in Table 2), even though no unit shows a cost efficiency of 100%. More generally, these approaches rank in Table 2 the units in the same way in terms of overall cost efficiency. However, the units are not ranked the same way on each component of overall cost efficiency by all the approaches. For example, the TT approach identifies unit E as the least price efficient (i.e. with largest scope for savings by adjusting input prices), whereas under the CD approaches and our (PT) approach unit C is the one showing lowest price efficiency or highest potential to reduce costs through price changes (which is an intuitive result as unit C is the unit showing highest prices of inputs). In addition, the TT model and our approach (PT) identify for unit E that the only way to reduce costs is by changing prices, while the CD approaches do identify cost savings due both to input mix and input price changes, even if target quantities for unit E in the second model of the CD (i) approach are equal to observed input quantities.

We shall use unit B to illustrate why our approach provides a clearer decomposition of cost savings than the CD and TT approaches. Consider first the CD(ii) approach, where for unit B, results in Table 2 show that savings of nearly 70% (down to 30.77%) of observed aggregate cost of inputs can be achieved. These savings are disaggregated into the following components:

- Reducing costs down to 38.46% of observed level and this is derived through the Farrell cost efficiency model (1). These cost savings are achieved by changing input quantities from (x_1, x_2)= (3, 2) to (1, 2), which is the optimal combination of inputs under the observed prices (4, 0.5) (see results of FM model in Table 2) ;

- Reducing the Farrell cost value obtained from (1) further down to 80% of its

18

value. This can be obtained by assuming that the input prices of unit D can be adopted. Thus these savings are achieved if prices change from (4, 0.5) to (1, 2). But the model solved for these prices identifies also optimal quantities, which are (2,1).

These optimal quantities are, however, disregarded in the CD (ii) approach (as well as in the CD (i)) as all cost changes from the Farrell minimum cost to the overall minimum cost are interpreted as market efficiency. Therefore there is no component of the cost efficiency measure of the CD approaches that encapsulates changes in the quantity of inputs required to achieve minimum aggregate cost of inputs. This approach, however, did identify a component of inefficiency for unit B associated with input price changes which the TT approach failed to identify.

Indeed, the TT approach identifies for unit B only two sources of inefficiency (technical and allocative) and considers it price efficient. Technical efficiency is identified in the TT approach through a traditional technical efficiency model, and its value is 60% (see Figure 1), meaning that the unit can save 40% of actual costs if it reduces input quantities pro rata to 60% of observed levels. Price and allocative efficiency are defined in a cost space (as defined in Figure 2), where inputs costs (of technical efficient quantities) are represented on the axis. Units that are on the 'technical' frontier of this cost space representation are deemed price efficient (and such is the case of unit B). However, at the point where unit B is located costs are not minimum and it would need to change the costs from $(C_1, C_2) = $ (7.2, 0.6) to (1.8, 2.4) to become efficient in terms of cost. These changes imply a change in costs to 53.85% of technically efficient costs. The latter are achieved by changing the mix of costs from point B to point D. Note, however, that movements from B to D (which under the TT approach are called allocative efficiency) can be achieved either by changing prices and/or changing input quantities or both (as we are in a cost space cost savings can be achieved by savings in quantities and/or prices). Yet these changes are not disentangled in the TT model. Clearly, if reduction in costs, via changing the mix of technically efficient costs, is achieved through a change in prices the allocative efficiency will indeed reflect price efficiency. Therefore the true price efficiency is not entirely captured by the price efficiency component of the TT approach.

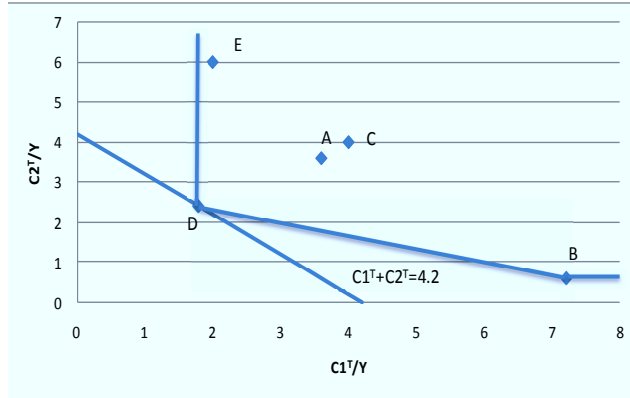Our proposed (PT) decomposition, while identifying the same minimum aggre-

Figure 2: Cost production technology of Tone and Tsutsui (2007) for illustrative example.

gate input cost as the CD(i) model, provides a decomposition that clearly disentangles savings that accrue from changing quantities and savings that accrue from changing prices. It suggests that to be overall cost minimizing unit B should change its input quantities from (3, 2) to (1, 2) and its input prices from (4, 0.5) to (1, 0.5). Therefore overall cost savings of 84.62% of observed cost can be partitioned into 38.46% obtained from changing input quantities and 46.15% from changing input prices.

The example in Table 2 demonstrates how our approach can capture more clearly both the scope for efficiency savings and its decomposition more accurately than existing approaches. Under the PT approach cost savings can be negative, as it happens for unit C, meaning that quantity targets in fact imply a rise in costs when assessed at average prices. This rise is however compensated for by the cost gains obtained from changing prices so that overall the unit can save 75% of its aggregate costs by changing quantities and prices together.

We show in Table 3 the decomposition of the potential quantity gains and price gains into a radial component and a mix component for the PT model, to illustrate the advantages of pursuing such a decomposition.

In Table 3 it is clear that for units A, B, C, and E the potential savings due to quantity changes can be obtained by changing the mix of inputs rather than changing

Table 3: Price and quantity cost gains decomposed into radial and mix components

	Total quantity	radial quantity	mix quantity	Total price	radial price	mix price
A	25.00%	0.00%	25.00%	58.33%	41.67%	16.67%
B	38.46%	0.00%	38.46%	46.15%	0.00%	46.15%
C	-9.38%	0.00%	-9.38%	84.38%	56.25%	28.13%
D	53.57%	39.29%	14.29%	32.14%	0.00%	32.14%
E	0.00%	0.00%	0.00%	75.00%	50.00%	25.00%

them radially. On the contrary, savings for unit D can be obtained by changing both the input quantities pro-rata and their mix. Unit C is the one showing the highest impact on cost due to changes in the mix of input quantities (since its observed cost increases by a percentage of 9.38%). Note that this unit has observed quantities of (2,1) and targets under the PT approach are (1,2), which means a complete reversal in the mix of quantities of inputs used to secure the output level of this unit.

Cost gains due to price changes, in Table 3 happen for units B and D only as a result of mix changes, and for the remaining units as a result of both radial and mix price changes. Unit B shows the highest cost gain accruing from a change in the mix in prices, since observed prices for this unit are (4, 0.5) and optimum prices are (1, 0.5). These mix changes reflect solely the need for this unit reducing the price of input 1, while keeping the price at which input 2 is secured.

In summary, the above shows that the decomposition we propose is more appropriate than those found in existing approaches dealing with savings attributable to input price and quantity changes when units are not price takers. The decomposition proposed in this paper reflects the fact that when units are not price takers price and quantity optimization are simultaneous. Therefore it does not make sense to optimise quantities for observed prices as done in other approaches and thereby deduce savings due to quantity changes and deduce residually 'allocative' efficiency. Optimal quantities at observed prices may not be optimal for optimal prices and allocative efficiency as traditionally defined makes no sense when units are not price takers. In addition, our decomposition still allows the differentiation between cost gains that accrue from reducing quantities and prices in a radial way or in a way that mainly implies a change in the mix of quantities and/or prices.

5.2 Modelling price relationships

In the previous section we showed that our proposed model can provide a clear decomposition of total scope for savings into those attributable to quantity and to price changes. The model developed is however a general one, which put forth for the first time the implicit mathematical modelling of price relationships of some existing models (like those of Camanho and Dyson (2008)), opening the avenue for further price modellization.

In this section, we show how our model can identify the scope for savings through the introduction of additional constraints on prices that may avoid target prices that are not reachable or which are unreasonable. For example, under the CD (ii) approach some units need to raise some input prices to be price efficient (e.g. in table 2 unit B should reduce the price of input 1 from 4 to 1, but increase the price of input 2 from 0.5 to 2). This is clearly counter intuitive from a perspective of cost minimisation. This situation arises because the CD(ii) approach considers an intensity variable (z_j) associated to each unit j rather than intensity variables associated to each input i. This is equivalent to consider the existence of trade-offs between input prices. In some circumstances this can indeed be the fact (e.g. if 2 inputs are bought from the same supplier, a reduction in the price of one of them may imply an increase in the price of the other), but one cannot expect that substantial increases in factor prices are reasonable in a context of cost minimisation. Even if one accepts the existence of such trade-offs one may not accept price increases, and therefore a suitable modification to model (3), when z_{ij} are replaced by z_j, is to impose further constraints on price targets to avoid them to increase ($\gamma_i \leq 1$).

Additional restrictions on factor prices can also be generalized to situations where price and quantities are linked in some way (as in the case of price discounts for large quantities bought from a supplier). For example, assume that the price of a certain input i cannot decrease below observed price (p_{io}) unless the company orders a quantity higher than Q, in which case a quantity discounts of k% will apply. In order to model this situation in model (3) one could add the following constraints: $p_i = p_{io} - k\%p_{io}b_i$, and $x_i \geq Qb_i$, where b_i is a binary variable (note that $p_i = \gamma_i p_{io}$). Such flexibility in capturing restrictions on input prices and linkages between prices and quantities are not available in existing approaches addressing the issue of cost efficiency under varying input prices.

To illustrate how our model can cope with additional price constraints, we have chosen to solve model (3) imposing convexity on factor prices (i.e. $z_{ij} = z_j$ for all i) and simultaneously imposing limits on factor prices, which cannot increase above observed levels and cannot decrease below 80% of observed levels (this level is arbitrary, but is used here for illustrative purposes only). Such additional constraints take the form: $\alpha_i \leq \gamma_i \leq 1$ in model (3). The results from this modified model are shown in Table 4.

Table 4: Results from model (3) with $\alpha_1 = \alpha_2 = =0.8$ and $z_{ij} = z_j$ for all i

	A	B	C	D	E
Obs cost	12	13	8	14	8
Min cost	5.6	5	6.4	4	5.6
Target (x1, x2)	(2, 1)	(1, 2)	(2, 1)	(2, 1)	(2, 1)
θ_1	0.667	0.333	1	0.333	2
θ_2	0.5	1	1	0.25	0.5
Target (p1, p2)	(1.6, 2.4)	(4, 0.5)	(1.6, 3.2)	(1,2)	(1.6, 2.4)
γ_1	0.8	1	0.8	1	0.8
γ_2	0.8	1	0.8	1	0.8
Potential savings due to input quantity changes	37.50%	61.54%	0.00%	71.43%	11.25%
Potential savings due to input price changes	15.83%	0.00%	20.00%	0.00%	18.75%
Total potential savings	53.33%	61.54%	20.00%	71.43%	30.00%
Cost efficiency	46.67%	38.46%	80.00%	28.57%	70.00%

With these additional constraints two units show a price efficiency of 100% (B and D), while under the CD(ii) approach only unit D would be 100% price efficient. The constraints imposing limits on price changes resulted in several units being required to reduce their input prices down to 80% of observed levels (A, C, and E), as seen in the optimal γ_i^* factors. Note however, that this reduction does not translate necessarily in a percentage cost gain of 20%, unless the unit is efficient in quantity terms (see unit C).

The above example shows that our approach can identify total savings that are more realistic than existing approaches, by limiting the variations that can indeed happen on input prices. Results in Table 4 identify higher cost efficiency values for all the units, based on a more reasonable assumption regarding factor price changes.

6   Conclusion

In a considerable number of cases in real life, markets are not perfectly competitive and DMUs may have some degree of control over prices.This paper has addressed the issue of computing and decomposing cost efficiency when DMUs being assessed are not strictly input price takers. It is assumed DMUs could gain by manipulating simultaneously, to the degree possible, input prices and quantities. The approach proposed puts forth for the first time a DEA model that optimizes prices and quantities simultaneously. This reflects more appropriately than previous approaches the fact that the units being assessed can have scope to decide and therefore perform well or poorly both on the choice of input/output quantities and on exploiting such flexibility as there may be on unit prices. Further, the model proposed is a general model that can be modified to include different types of restrictions on prices, to capture local conditions faced by a unit such as bounds on price changes or links between individual input prices and quantities.

The paper offers a suitable decomposition of the potential cost reductions between those attributable to potential input quantity adjustments and those attributable to input price adjustments. It is argued within this paper that when prices are not fixed the traditional concept of allocative efficiency loses its meaning. Therefore the decomposition to be made on total cost savings pertains to the savings that can be achieved through quantity changes and the savings that can be achieved through price changes. No existing approach in the literature isolates these two components so clearly when both quantities and their prices can vary simultaneously.

References

Athanassopoulos A and Gounaris C. 2001. Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications. *European Journal of Operational Research* **133** 416–431.

Balk, B.M., Färe, R, and Grosskopf, S. 2004. The theory of economic price and quantity indicators. *Economic Theory* **23** 149–164.

Camanho, AS. and Dyson, R. 2005. Cost efficiency measurement with price un-

certainty: a DEA application to bank branch assessment. *European Journal of Operational Research* **161** 432–446.

Camanho, AS. and Dyson, R. 2008. A generalisation of the Farrell cost efficiency measure applicable to non-fully competitive settings. *Omega* **36** 147–162.

Charnes A, Cooper WW, Rhodes E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* **8/2** 429–44.

Chambers, RG. 2002. Exact nonradial input, output, and productivity measurement. *Economic Theory* **20** 751–765.

Cherchye, L, Kuosmanen, T and Post, T. 2002. Non-parametric production analysis in non-competitive environments. *Int. J. of Production Economics* **80** 279–294.

Diewert , WE. 2005. Index number theory using differences rather than ratios. *The American Journal of Economics and Sociology* **64/1** 311–360.

Farrell, MJ. 1957. The measurement of productive efficiency. *Journal of Royal Statistical Society Series A* **120** 253–381.

Färe, R , Grosskopf, S and Nelson, J. 1990. On price efficiency. *International Economic Review* **31/3** 709–720.

Färe, R, Grosskopf, S and Lovell, CAK.1994. Production Frontiers. Cambridge University Press.

Färe R, Grosskopf S and Lovell CAK. 1985. The measurement of efficiency of production. Boston: Kluwer Academic.

Färe, R , and Grosskopf, S. 2006. Resolving a strange case of efficiency. *Journal of the Operational Research Society* **57** 1366–1368.

Grifel-Tatjé, E and Lovell, CAK. 1999. Profits and Productivity.*Management Science* **45/9** 1177–1193.

Grifel-Tatjé, E and Lovell, CAK. 2000. Cost and Productivity.*Managerial and Decision Economics* **21** 19–30.

Greene, WH. *The Econometric approach to efficiency analysis.* In Fried, H. O. , Lovell, C. A. K. and Schmidt, S. S., Eds. The measurement of productive efficiency and productivity growth, Oxford University Press. 92-250.

Johnson, A L, Ruggiero, J. 20011. Allocative efficiency measurement with endogenous prices . *Economics Letters* **111** 81–83.

Kuosmanen, T and Post, T. 2002. Nonparametric efficiency analysis under price uncertainty: a first order sytochastic dominance approach. *Journal of Productivity Analysis* **17** 183–200.

Kuosmanen, T, Kortelainen, M, Sipilainen, T, Cherchye, L. 2010. Firm and industry level profit efficiency analysis using absolute and uniform shadow prices. *European Journal of Operational Research* **202** 584–594.

Hervé Leleu. 2013. Inner and outer approximations of technology: A shadow profit approach. *Omega* **41** 868–871.

Pintér, JD 2007. Nonlinear Optimization with GAMS/LGO.*Journal of Global Optimization* **38/1** 70–101.

Sahoo, B.K. and Tone K. 2013. Non-parametric measurement of economies of scale and scope in non-competitive environment with price uncertainty. *Omega* **41** 97–111.

Ray, S, Chen, L and Mukherjee, K. 2008. Input price variation across locations and a generalised measure of cost efficiency.*International Journal of Production Economics* **116** 208-218.

Russell, RR. 1985. Measures of technical efficiency.*Journal of Economic Theory* **35/1** 109-126.

Thanassoulis, E, Portela, MCAS, and Graveney, M. 2012. Estimating the scope for savings in referrals and drug prescription costs in the General Practice units of a UK Primary Care Trust. *European Journal of Operational Research* **221** 432-444.

Thanassoulis, E, Portela, MCAS, and Despic, O. 2008. Data Envelopment Analysis: The Mathematical programming Approach to efficiency Analysis. In Fried, HO,

Lovell, CAK, and Schmidt, SS, The measurement of Productive Efficiency and Productivity Growth. Oxford. 251-420.

Tone K. 2002. A strange case of the cost and allocative efficiencies in DEA. *Journal of the Operational Research Society* **53** 1225–1231.

Tone, K, and Tsutsui, M. 2007. Decomposition of cost efficiency and application to Japanese-US utility comparisons.*Socio-Economic Planning Sciences* **41** 91–106.