CHAPTER 1

INTRODUCTION

1.1 Motivation

Firms generate something with exchange value. This value may be either tangible or intangible. The principal characteristic of this generation is the transformation of inputs into outputs. Inputs may be knowledge, labour, capital, raw materials, intermediate goods, services, etc. Outputs may be services, intermediate goods, finished goods, etc. The generation of outputs is production.

Production is made through one or more processes. In a competitive economic system, the way a firm defines and executes its processes determines the economic result of the firm, because the quantities of inputs that are needed for producing an output depend, not only on the output quantities, but also on the technical and economic efficiency of the production processes.

For a particular process, in terms of products, technical efficiency is the ratio of the actual production quantity to the maximum quantity that is feasible. The actual and the maximum production quantities are for the same quantities of the inputs. Correspondingly, in terms of inputs, technical efficiency of a production process is the ratio of the minimum required quantities of inputs to the actual quantities of inputs. The actual and the minimum input quantities are for the same quantities of outputs

Although it is not always feasible to know the theoretical minimum quantities of inputs or the maximum quantities of outputs, it is always possible to know the values of the inputs and outputs for the best-performing firms in the industry. These "actual best quantities" are valid as benchmarking values. Best-performing firms are the most technically efficient firms.

Non-efficient firms must adjust the quantities and mix of inputs and/or outputs to perform as fully efficient. *Targets* are the quantities of inputs and outputs that a non-efficient firm has to match, to perform as fully efficient. Targets are such that they optimise some objective function of the firm.

After target definition, the problem that arises is to determine the optimal sequence of adjustments to the inputs and/or outputs, which period-to-period non-efficient firms must match in order to achieve the targets that optimise the objective of the firm. To solve this problem is the main motivation of this research.

1.2 Objectives and Main Results of this Study

As stated in Section 1.1, the purpose of this study is the design of optimal paths of adjustment of inputs, from the actual input quantity vector, $\boldsymbol{x}$, to the target input quantity vector, $\boldsymbol{x}^*$, keeping output quantity vector, $\boldsymbol{y}$, constant within each adjustment period. The path of adjustment is optimal in the sense that it minimises the present value of the cost of period-to-period inputs and the present value of period-to-period costs of adjustment of inputs. Orientated to this purpose, this section describes the optimising behaviour of firms.

The main objective of this research is to extend the basic Data Envelopment Analysis (DEA) methodology in the following areas:

(a) *To assign the target that maximises profit.* Profit maximisation requires knowledge of input and output prices. For input-orientated systems with given constant outputs, profit maximisation and minimisation of the total cost of inputs give the same optimal input quantity vectors. Prices may or may not be the same for all firms. For example, labour costs may differ from firm to firm. Equal input prices for different firms is not a necessary condition in this research. The reason for this is that the maximum profit is determined separately for each non-efficient firm. The targets have to be feasible from a physical, technical, legal, and economic point of view. The targets also must match 100 per cent technical and economic efficiency, as measured using DEA methodology.

(b) *To select the optimal path of adjustment.* Once the targets are defined, the optimal path of adjustment is determined. The optimal path maximises the firm's present value of the cash flow of sales minus input costs and the cost of adjusting the input quantity vectors or increasing the output quantity vectors. The costs of adjustment must be known so that the optimal path of adjustment can be determined.

For the purposes of this research, the following considerations apply:

(a) The evaluation of the present value of cash flows is at a constant compound rate of discount.

(b) For input-orientated systems, the evaluation of the present value of cash flows is on the assumption that adjustments are done at the start of each period. This means that the cost of adjusting the input quantity vector is incurred at the start of the period and that the savings (or increased costs) are perceived at the end of it.

The methodology of this research is applied to Dijon, a chain of Chilean department retailing stores. Dijon is orientated to middle-income people, and has 57 stores in Chile. A description of Dijon is in Section 6.2 of Chapter 6.

The two main results of the research are:

1) It is possible to determine optimal paths of adjustment of inputs, from the initial input quantity vector to the target one. The optimal path of adjustment is strongly conditioned by the period-to-period budget constraints.

2) The weights of firms may differ from period-to-period. This consideration allows for period-specific description of technology and the determination of the period-specific optimal input quantity vector. The period-specific description of technology may include period-to-period changes in the technology or dynamic DEA. A period-specific description of the technology may include the use of different technologies.

1.3 Thesis Outline

This section outlines the contents of the subsequent chapters and sections of the thesis.

Chapter 2 presents basic concepts of production functions, productivity and efficiency measurement. Section 2.2 introduces the production function of firms that produce one output using one or more inputs. Section 2.3 extends the analysis to the transformation function of firms that produce two or more outputs using two or more inputs. Section 2.4 presents maximisation of profit, minimisation of cost of inputs, and maximisation of revenue as expected behavioural motivations for producers. Section 2.5 introduces the concepts of technical, allocative, and economic efficiency. Technical efficiency is the starting concept for the development of the dual DEA model in the next chapter. Section 2.6 introduces the concept of productivity. This is the starting concept for the development of the primal DEA model in Chapter 3. Section 2.7 presents two basic models for the analysis of the dynamics of efficiency. The concepts associated with dynamics of efficiency, together with the dual DEA model, are the bedrock of Chapter 4, where the basic model of optimal paths of adjustment is introduced. Section 2.8 presents extensions of the basic concepts of the dynamics of efficiency to primal and dual models of dynamic production systems. Section 2.9 closes Chapter 2 with some conclusions.

Chapter 3 presents fundamental concepts of Data Envelopment Analysis. Section 3.2 states the primal DEA problem, and discusses productivity as the largest value of a weighted sum of outputs divided by a weighted sum of inputs. The weights are the linear programming optimisation variables that are found. A simple example illustrates these concepts. Section 3.2 closes with a brief presentation and interpretation of the information obtained from the solution of that example. Section 3.3 states the dual DEA problem, and relates it with technical efficiency. Section 3.4 presents, in general terms, the concept of returns to scale. Section 3.5 reviews, in DEA formulation, the expected optimising behaviour of firms. Section 3.6 discusses some limitations of DEA. Section 3.7 presents some concluding comments.

Chapter 4 presents a basic dual DEA model for the selection of optimal paths of adjustment. This basic dual DEA model is derived from the fundamental dual DEA problem presented in Chapter 3. Section 4.2 introduces concepts used in this and following chapters: adjustment, cost of adjustment, dynamic DEA, intended number of adjustment periods, input target, path of adjustment, period, present worth value factor, profit, time of adjustment, and time horizon. Section 4.3 presents a basic optimal control model for the selection of optimal paths of adjustment. This model is referred to as the basic dual DEA model. The basic dual DEA model introduces the concept of dynamic DEA as a sequence of period-to-period static dual DEA problems. Period-to-period adjustment of the input quantity vector has a corresponding period-to-period specific vector of weights of peers. Peers may change weight from period to period and peer firms may be different from one period to the next. Section 4.4 presents an example of the application of the basic

DEA model. Section 4.5 presents some conclusions from the application of the basic DEA model. Section 4.6 closes this Chapter with Conclusions.

Without loss of generality, in the following chapters of this thesis we refer mainly to input-orientated systems. We extend the application to output-orientated systems with the use of examples.

Chapter 5 extends the basic model of Chapter 4. The basic model considers symmetric costs of adjustment of inputs. This means that the cost of increasing and the cost of decreasing inputs are the same. Section 5.2 presents the model of optimal paths of adjustment with asymmetric costs of adjustment. The basic model has the implicit assumption that the time of adjustment and the horizon time are the same. For cash-flow evaluation purposes, Section 5.3 permits the time horizon to be longer than the time of adjustment. Section 5.4 presents an example that includes asymmetric costs of adjustment and that the time horizon is longer than the time of adjustment. Section 5.5 presents the model of optimal paths of adjustment with dynamic (time-variable) outputs. Section 5.6 presents an example with dynamic outputs, asymmetric costs of adjustment, input prices, output prices and costs of adjustment variables. Section 5.7 presents the incorporation of quasi-fixed (non-discretionary) variables and an example. Section 5.8 presents the incorporation of a capital investment constraint. Section 5.9 presents the general model of optimal paths of adjustment with a dynamic boundary of the technology. Without loss of generality, the improved models consider input-orientated systems and constant returns to scale. The output-orientated systems and variable returns to scale are trivial extensions. Section 5.10 presents conclusions from the extensions to the basic model for selection of optimal paths of adjustment.

Chapter 6 presents a realistic application of the models of optimal path of adjustment, developed in Chapter 4 and extended in Chapter 5. The application uses relevant data of 35 comparable retail-level stores. The stores are branches of Dijon, a Chilean retailing firm. Section 6.2 is a general overview of Dijon. Section 6.3 presents the relevant data to be considered in this application. Because of a confidentiality agreement with Dijon, the data are encrypted by a simple escalation procedure. Section 6.4 presents the optimal path of adjustment LP model for Dijon's stores. Section 6.5 presents the optimal paths of adjustment for five stores, designed with the LP model. Section 6.6 presents the main economic results of optimal paths of adjustment designed with the LP model. Section 6.7 presents the main implication of considering period-to-period changes of weights of stores. Section 6.8 presents the case for Store 202 and its optimal paths of adjustment for expected period-to-period outputs, with variable prices and costs of adjustment. Section 6.9 presents the case for Store 202 with its optimal output quantity vectors and optimal paths of adjustment, considering expected period-to-period variable prices and costs of adjustment. Finally, Section 6.10 presents conclusions derived from the application of the optimal path of adjustment LP model to the 35 stores.

Chapter 7 presents a summary of the conclusions derived from the application of the optimal path of adjustment LP and opportunities for further research.

CHAPTER 2

PRODUCTION FUNCTION, PRODUCTIVITY AND EFFICIENCY MEASUREMENT

2.1 Introduction

This chapter presents the basic concepts of a production function, productivity and efficiency measurement. Section 2.2 introduces a production function for firms that produce one output using one or more inputs, and Section 2.3 extends the analysis to a transformation function of firms that produce two or more outputs using two or more inputs. Section 2.4 presents maximisation of profit, minimisation of cost of inputs, and maximisation of revenue as expected behavioural motivation for producers. Section 2.5 introduces the concepts of technical, allocative, and economic efficiency. Technical efficiency is the starting concept for the development of the dual DEA model in the next chapter. Section 2.6 introduces the concept of productivity, which is foundational for the development of the primal DEA model in Chapter 3. Section 2.7 presents two basic models for the analysis of the dynamics of efficiency. The concepts associated with the dynamics of efficiency, together with the dual DEA model, are foundational to Chapter 4, where the basic model of optimal paths of adjustment is introduced. Section 2.8 presents extensions to the concepts of dynamics of efficiency to primal and dual models of dynamic production systems. Section 2.9 closes this chapter with some conclusions.

2.2 The Production Function

As mentioned in Chapter 1, firms generate something with exchange value that is tangible or intangible. This generation is through the transformation of inputs into outputs. Inputs are knowledge, labour, capital, raw materials, services, utilities, intermediate goods, etc. Thus, senior employees supply experience (a crucial form of knowledge); blue- and white-collar workers supply labour; nature provides raw materials; and departments provide services such as data processing, accounting or quality control. Utilities use steam, refrigeration, or electric energy. Other firms or other sections of the same firm provide intermediate goods such as doorframes and boards for assembling doors, or tyres for a car-assembly plant.

Outputs may be services, utilities, finished goods, intermediate goods for the use of the same firm, etc. Production is the generation of outputs using inputs. A production process is a set of organised, repetitive, and coordinated tasks that transforms inputs into outputs.

The production technology is the ability of a production process to transform inputs into outputs. Production technology is frequently referred to simply as technology. Hereafter, both terms are used interchangeably.

A basic assumption is that there exists a relationship between inputs and outputs that can be written in mathematical form (Chambers, 1988, p. 7). The production function is the functional expression for one output in terms of one or more inputs, and the transformation function is the functional expression for more than one output in terms of one or more inputs.

For a production process that produces one output by using one or more inputs, the *production function* expresses the maximum output quantity attainable from given input quantity vector. The functional expression for the production function is:

$$y^* = \mathrm{f}(\boldsymbol{x}), \qquad (2.2.1)$$

where y^* is a scalar that represents the maximum quantity of a non-negative single output that is attainable from a non-negative input quantity, $\boldsymbol{x}$.

The most commonly used mathematical form for $\mathrm{f}(\boldsymbol{x})$ is the Cobb-Douglas production function. For one output and I inputs, the function is:

$$y^* = \alpha_0 \prod_{i=1}^{I} x_i^{\alpha_i}, \qquad (2.2.2a)$$

where x_i is the *i-th* input quantity and the α_is are unknown parameters to be estimated. The logarithmic form of equation (2.2.2a), given in equation (2.2.2.b), is preferred for the estimation of the α-parameters because it is linear in terms of the parameters and is easier to handle:

$$\log(y^*) = \log(\alpha_0) + \sum_{i=1}^{I} \alpha_i \log(x_i). \qquad (2.2.2b)$$

The α-parameters are frequently estimated using econometric methods. Equation (2.2.3) presents a generalised version of the Cobb-Douglas production function (Chambers, 1988, p. 27). If $f_i(\boldsymbol{x}) = \alpha_i$, and $g(\boldsymbol{x}) = 0$, Equation (2.2.3) reduces to the standard Cobb-Douglas production function.

$$y^* = \alpha_0 \prod_{i=1}^{I} x_i^{f_i(x)} \, e^{g(x)}. \qquad (2.2.3)$$

If $f_i(\boldsymbol{x}) = \alpha_i$, and $g(\boldsymbol{x}) = \sum_{i=1}^{I} \gamma_i x_i$, Equation (2.2.3) represents the *transcendental function*, where the γ_i s are unknown parameters to be estimated.

One of the assumptions of the Cobb-Douglas model is that the average cost of production first decreases and then increases monotonically with increasing output. Zellner and Revankar (1969) proposed a generalisation of the Cobb-Douglas production function, relaxing that assumption.

Alternative nonparametric methods for fixing the boundary of the production technology may be used. Chapter 3 introduces Data Envelopment Analysis (DEA) as the nonparametric method that is used in the rest of this thesis.

Figure 2.1 presents the production technology for the simplest case: the one-input and one-output process. The curve, OM, defines the boundary of the technology. The feasible set of output quantities is below this line and above y = 0. This means that the line OM defines the maximum quantity of the output that is feasible to produce for a given input quantity.

Following Chambers, (1988, pp. 21-24), the *average output* and the *marginal product* are defined as y/x and dy/dx respectively. The average output is the quantity of output per unit of input. For example, in Figure 2.1, the average output is the slope of the ray from the origin to the point of interest, B, with coordinates (y_e, x_e). The average output at (y_e, x_e) is y_e/x_e. In Figure 2.1, the maximum average output occurs at the point D that has coordinates (y_1, x_1).

The marginal output is the rate of change of output per unit change of input. In Figure 2.1, the marginal output is the slope of the line TT′ that is tangent to the curve at the point of interest (y_e, x_e). Then, the marginal output at (y_e, x_e) is $\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{y_e, x_e}$.

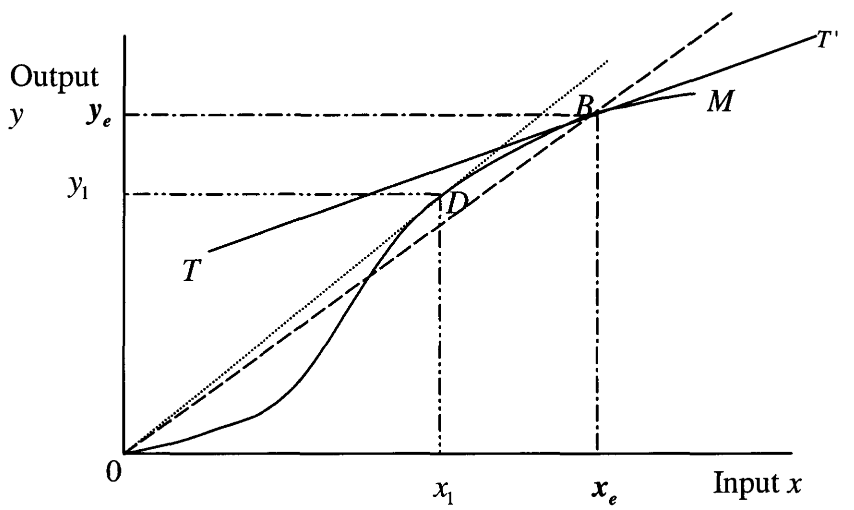

Figure 2.1: Representation of a Single-Output and Single-Input Production Technology

For this simple production function, the *elasticity of scale*, ε, is defined by:

$$\varepsilon = \left.\frac{x}{y}\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{y_e, x_e} = \left.\frac{\mathrm{d}log(y)}{\mathrm{d}log(x)}\right|_{y_e, x_e}.$$

If ε exceeds unity, then the production function exhibits *increasing returns to scale.* This means that, in the neighbourhood of x_e, increasing x results in increasing y in a larger proportion than x. If ε is equal to unity, then the production function exhibits *constant returns to scale*. This means that, in the neighbourhood of x_e, increasing x results in increasing y in the same proportion as x. This is the case of point D in Figure 2.1, because ray OD is at a tangent to the curve. Finally, if ε is less than unity, then the production function exhibits decreasing returns to scale, as at point B. This means that, in the neighbourhood of x_e, increasing x results in increasing y in a smaller proportion than x.

From this simple case it may be concluded that returns to scale is a local property of the technology. Section 3.4 of Chapter 3 presents the DEA criteria for knowing if the returns to scale of a technology are constant or variable, in a neighbourhood.

Figure 2.2 presents the production technology for a two-input and one-output process. For this process, the representation is in terms of isoquants. An isoquant is the locus of all combinations of quantities of inputs that yield the same output quantity. At any point, the slope of the isoquant is the marginal rate of technical substitution. The marginal rate of technical substitution is the rate at which one input substitutes some quantity of the other, while the output quantity remains constant. For a fixed output quantity, the substitution of some quantity of one input for some quantity of other inputs is fundamental for the purpose of this thesis, because the paths of adjustment involve input substitution in a non-radial modification of inputs.

Figure 2.2 presents the isoquant for y_1 and for y_2 units of output, such that $y_2 > y_1$.

Chambers (1988, p. 9) presents ten properties that the functional relationship (2.2.1) between inputs and output must fulfil to be a valid production function. For the purpose of this thesis, the following four properties are of interest:

PF1 monotonicity: if $\boldsymbol{x}' \geq \boldsymbol{x}$ then $(y' = f(\boldsymbol{x}')) \geq (y = f(\boldsymbol{x}))$;

PF2 convexity: $f(\theta \boldsymbol{x}^0 + (1-\theta)\boldsymbol{x}') \geq [\theta f(\boldsymbol{x}^0) + (1-\theta) f(\boldsymbol{x}')]$ for $0 \leq \theta \leq 1$;

PF3 weak essentiality: $f(0_i) = 0$, where 0_i is a null vector; and

PF4 $f(\boldsymbol{x})$ is finite, non-negative, real valued, and single valued for all non-negative and finite, $\boldsymbol{x}$.

The monotonicity property, PF1, implies that, when input quantities are increased, output will never decrease. The convexity property, PF2, implies that the law of diminishing marginal rate of technical substitution and the law of diminishing marginal productivity are satisfied. The law of diminishing marginal productivity states that beyond a certain point of production, increasing the quantity of specific input results in smaller and smaller increases in output. Mathematically, this means that the second-order partial derivatives of output with respect to inputs are increasingly negative. The weak essentiality property, PF3, states that, in the absence of all inputs, no output is produced. Nonetheless, only one or more inputs generate output.

To be a valid boundary of technology, these four properties must also be satisfied by the sectionally-linear approximations to a production function that is implicit in DEA. As was stated before, Chapter 3 presents some fundamental concepts of DEA.

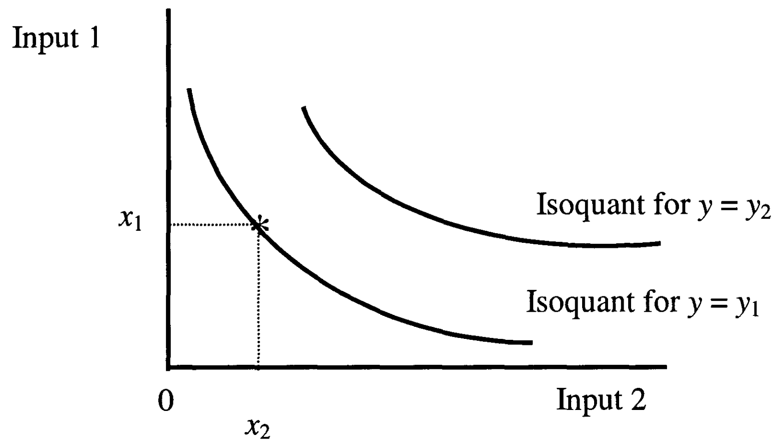

Figure 2.2: Representation of a One-output, Two-input Production Technology

2.3 The Transformation Function

In this section, the multiple-input and single-output case is generalised to the multiple-input and multiple-output technology.

The transformation function is the mathematical expression for the technological relationship between multiple inputs and multiple outputs. For a process with I inputs and J outputs, the general form of the transformation function is:

$$T(\boldsymbol{x}, \boldsymbol{y}) = 0, \tag{2.3.1}$$

where $\boldsymbol{x}$ is a $(I \times 1)$ input quantity vector and $\boldsymbol{y}$ is a $(J \times 1)$ output quantity vector. For a specific process, equation (2.3.1) gives the minimum $\boldsymbol{x}^*$ that is technologically feasible to use for producing a specified output quantity vector $\boldsymbol{y}$. Equivalently, for a specific process, equation (2.3.1) gives the maximum y^* that is technologically feasible to obtain using the input quantity vector $\boldsymbol{x}$. If y^* is a scalar, then equation (2.3.1) reduces to the production function that is presented in equation (2.2.1). Equation (2.3.2) presents the reduction of the transformation function to the production function.

$$T(\boldsymbol{x}, y^*) = y^* - f(\boldsymbol{x}) = 0. \tag{2.3.2}$$

Following Chambers (1988, pp. 252-261), consider the set of all combinations of inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$ that are technically feasible, for a given technology $T(\boldsymbol{x}, \boldsymbol{y}) \leq 0$. That set is referred to as the *production possibility set* and is represented by S. To be a valid production possibility set, S must fulfil the following properties:

PPS1 S is nonempty;

PPS2 S is a closed set;

PPS3 S is a convex set;

PPS4 if $(\boldsymbol{x}, \boldsymbol{y}) \in S, \boldsymbol{x}^{\bullet} \geq \boldsymbol{x}$, then $(\boldsymbol{x}^{\bullet}, \boldsymbol{y}) \in S$;

PPS5 if $(\boldsymbol{x}, \boldsymbol{y}) \in S, \boldsymbol{y}^{\bullet} \leq \boldsymbol{y}$, then $(\boldsymbol{x}, \boldsymbol{y}^{\bullet}) \in S$;

PPS6 for every finite $\boldsymbol{x}$, S is bounded from above; and

PPS7 $(\boldsymbol{x}, \boldsymbol{0}_j) \in S$, but if $\boldsymbol{y} \geq 0$, $(\boldsymbol{0}_i, \boldsymbol{y}) \notin S$.

Property PPS1 indicates that a technology exists. Property PPS2 says that all boundary points of S belong to S. Property PPS3 implies that any mixture of two feasible input quantity vectors is a feasible input quantity vector. Property PPS4 indicates that if input quantity vector $\boldsymbol{x}$ produces output quantity vector $\boldsymbol{y}$, then an increased input quantity vector produces no less than output quantity vector $\boldsymbol{y}$.

Property PPS5 indicates that if input quantity vector $\boldsymbol{x}$ produces the output quantity vector $\boldsymbol{y}$, then a decreased output quantity vector may be produced with the same input quantity vector $\boldsymbol{x}$. Property PPS6 guarantees the existence of a transformation function. Property PPS7 states that it is not possible to get something from nothing.

By definition, if $T(\boldsymbol{x}, \boldsymbol{y}) \leq 0$, then the input quantity vector $\boldsymbol{x}$ and the output quantity vector $\boldsymbol{y}$ are such that $(\boldsymbol{x}, \boldsymbol{y}) \in S$.

The convexity property of S requires that if $T(\boldsymbol{x}^0, \boldsymbol{y}^0) \leq 0$ and $T(\boldsymbol{x}^1, \boldsymbol{y}^1) \leq 0$, then $T[\{\theta \boldsymbol{x}^0 + (1-\theta)\ \boldsymbol{x}^1\}, \{\theta \boldsymbol{y}^0 + (1-\theta)\ \boldsymbol{y}^1\},\ 0 \leq \theta \leq 1] \leq 0$; this condition is satisfied if $T[\{\theta \boldsymbol{x}^0 + (1-\theta) \boldsymbol{x}^1\}, \{\theta \boldsymbol{y}^0 + (1-\theta) \boldsymbol{y}^1\}, 0 \leq \theta \leq 1] \leq \theta\, T(\boldsymbol{x}^0, \boldsymbol{y}^0) + (1-\theta) T(\boldsymbol{x}^1, \boldsymbol{y}^1)$

The PPS4 requires that if $T(\boldsymbol{x}, \boldsymbol{y}) \leq 0$, then $T(\boldsymbol{x}^1, \boldsymbol{y}) \leq 0$ for $\boldsymbol{x}^1 \geq \boldsymbol{x}$. If the input quantity vector $\boldsymbol{x}$ belongs to the technically efficient input set, then $T(\boldsymbol{x}, \boldsymbol{y}) = 0$.

Property PPS7 implies that $T(\mathbf{0}_i, \boldsymbol{y}) > 0$ for $\boldsymbol{y} \geq 0_j$.

These basic properties are equally valid for continuous transformation functions and for sectionally continuous functions. For the purpose of this thesis, Chapter 3 presents DEA as a non-parametric stepwise linear approximation to a transformation curve that is constructed using observed production data. The linear approximation of observed production data may involve straight lines, planes or hyperplanes, depending on the number of inputs and outputs involved.

The production function, as special case of the transformation function, fulfils the properties indicated above. PF1 (monotonicity) satisfies PPS4; PF2 (convexity) satisfies PPS3; PF3 (essentiality) satisfies PPS7; and PF4 (finite, non-negative, real valued, and single valued for all non-negative and finite $\boldsymbol{x}$) satisfies PPS6.

The transformation function $\boldsymbol{T}(\boldsymbol{x},\boldsymbol{y})=0$ gives the maximum output quantity vector that is possible to obtain using a specified input quantity vector. Also, the transformation function gives the minimum input quantity vector that is feasible to use for producing a specified output quantity vector.

2.4 Optimising Behaviour

As stated in Chapter 1, the purpose of this thesis is the design of optimal paths of adjustment of inputs, from an actual input quantity vector $\boldsymbol{x}$ to the target one $\boldsymbol{x}^*$, keeping output quantity vector $\boldsymbol{y}$ constant within each adjustment period. The path of adjustment is to be optimal in the sense that it minimises the present value of the total cost of inputs plus the costs of adjustment of inputs.

Orientated to this purpose, this section describes the three types of optimising behaviour of firms. In general terms, the behaviour of firms, under certainty, is profit maximisation (Chambers, 1988, p. 120).

For the purpose of this thesis, we use the accounting meaning of the term *net profit* as the residual after deduction of all money costs, i.e., sales revenues minus wages, salaries, rents, costs of raw materials, etc. (Bannock, Baxter and Davis, 1998, p. 335). The sales revenue is $\boldsymbol{yp}$, where $\boldsymbol{y}$ is the output quantity vector sold and $\boldsymbol{p}$ is the vector of output prices. The money cost is $\boldsymbol{xw}$, where $\boldsymbol{x}$ is the input quantity vector used to produce $\boldsymbol{y}$ and $\boldsymbol{w}$ is the vector of input prices. This optimisation assumes that prices $\boldsymbol{p}$ and $\boldsymbol{w}$ are known and independent of the quantities used. Following Chambers (1988, p. 54) and Doll and Orazem (1984, pp. 14-16), it is assumed that producers are atomistic competitors, in the sense that each one is small enough not to influence the market price in individual form; for this reason these producers take prices as given by the market. Producers in this situation are referred to as price takers. Also, it is assumed that there are no significant barriers to enter into or cease production of goods or services, and that there are no artificial restrictions on the supply or demand of those goods or services. These assumptions imply that prices are free to vary according to market equilibrium. If firms are considered as buyers, it is assumed that each individual firm does not affect the price when deciding to increase the quantity of some good it buys, or if it decides not to buy. This means that firms must regard prices of inputs as given.

Equation (2.4.1) describes the behaviour of the Firm e that maximises profit π_e :

$$\pi_e\,(\boldsymbol{p},\,\boldsymbol{w}) = \max_{\boldsymbol{x},\boldsymbol{y}}\ \{\pi\,(\boldsymbol{yp}\text{-}\boldsymbol{xw})\text{: } \mathrm{T}(\boldsymbol{x},\boldsymbol{y})=0\}, \qquad (2.4.1)$$

where $\pi\,(\cdot\ )$ is the profit function that captures the preference of the firm. Equation (2.4.1) states that firms select the set of technically feasible inputs and outputs that maximises profit. The optimisation is at given and known output and input prices. Input and output prices are exogenously determined.

Chambers (1988, p. 124) and Varian (1992, p. 49) state the following properties of the profit function:

PR1 Non-negativity: $\pi_e\,(\boldsymbol{p},\,\boldsymbol{w}) \geq 0$;

PR2 Non-decreasing in **p**: if $\mathbf{p}^1 \geq \mathbf{p}^2$, then $\pi_e(p^1, w) \geq \pi_e(p^2, w)$;

PR3 Non-increasing in **w**: if $\mathbf{w}^1 \geq \mathbf{w}^2$, then $\pi_e(p, w^2) \geq \pi_e(p, w^1)$;

PR4 $\pi_e(p, w)$ is convex and continuous in (p, w); and

PR5 Positivity and linear homogeneity: $\pi_e(\varsigma p, \varsigma w) = \varsigma\, \pi_e(p, w)$ for $\varsigma > 0$.

PR1 indicates that the firm will not operate, even in the short run, at negative variable profits. The firm must produce enough revenues to cover the cost of inputs. PR2 indicates that if output prices increase then profit will not decrease. PR3 indicates that if input prices increase then profit will not rise. PR4 is a curvature property derived from the convexity property, PF2, of the transformation function. PR5 indicates that, if all prices are changed by the same positive proportion, then the profit changes by the same proportion.

The profit function is usually used in industries where input and output markets are competitive and where producers are able to make rational decisions about the quantities of production, at competitive input and output prices.

Färe, Grosskopf and Lovell (1994, p. 213) point out that there may be no finite solution to the problem of equation (2.4.1). In fact, if no additional constraints are imposed to problem of equation (2.4.1) [and that of equation (3.5.1) in Chapter 3], there is no bounded solution for constant and increasing returns to scale technologies. The reason for this is that to maximise profit $\pi_e(p, w) = yp\text{-}xw$, y increases without bound (property PF1),. Unbounded y drives unbounded x. To avoid this situation, two separate cases are derived (Chambers, 1988, p. 121).

The first case is one of short-run optimisation that involves maximising profit for a given fixed output quantity vector $\boldsymbol{y}$. With the output quantity vector $\boldsymbol{y}$ constrained to be fixed, profit maximisation corresponds to a cost-minimisation problem. Consider the case of water distribution companies. In general terms, consumers have no substitute for water, and firms are not free to determine the price. To maximise profit, firms must minimise the cost of inputs. The cases of small farms and gas and electricity supply companies are similar.

Then, for the fixed output quantity vector $\boldsymbol{y}$, the main purpose is to determine the optimal input quantity vector, $\boldsymbol{x}^*$, that maximises the profit, π_e. This means that the optimal input quantity vector that maximises profit, $\pi_e\,(\boldsymbol{p}, \boldsymbol{w})$, is the same optimal input quantity vector that minimises the cost of inputs, $\boldsymbol{c(wx)}$. The profit-maximising problem becomes the cost-minimising problem:

$$c_e(\boldsymbol{wx}) = \underset{\boldsymbol{x}}{\text{minimise}}\,\{\boldsymbol{c(wx)}\text{: } \mathrm{T}(\boldsymbol{x}^*, y)=0\}. \qquad (2.4.2)$$

The optimal input quantity vector $\boldsymbol{x}$ is denoted by $\boldsymbol{x}^*$ to imply that it is on the boundary of the technology, described by $\mathrm{T}(\boldsymbol{x}^*, y)=0$.

In other words, the minimisation of the cost function gives the optimal input quantity vector, $\boldsymbol{x}^*$, that minimises the cost of producing a given output quantity vector, $\boldsymbol{y}$. The cost function is expressed as a function of the exogenous input prices and assumes that they are not zero. The cost function also depends on the transformation function, which specifies the inputs, $\boldsymbol{x}^*$, capable of producing the fixed output, $\boldsymbol{y}$.

Chambers (1988, p. 52) presents the following properties of the cost function, $c(w,y)$:

CO1 Non-negativity: $c(w,y) \geq 0$ for $w>0$ and $y>0$;

CO2 Non-decreasing in $\mathbf{w}$: if $\mathbf{w}^1 \geq \mathbf{w}^2$, then $c(w^1,y) \geq c(w^2,y)$;

CO3 $c(w,y)$ is concave and continuous in $\mathbf{w}$;

CO4 $c(w, y\ )$ is convex and continuous in y^*; and

CO5 Positivity and linear homogeneity: $c(\varsigma w, y) = \varsigma\ c(w, y)$ for $\varsigma > 0$.

These properties parallel those of the profit function.

The second case is one of long-run optimisation that involves maximising profit for a given fixed input quantity vector x. With the input quantity vector x constrained to be fixed, profit maximisation corresponds to a revenue-maximisation problem.

The revenue-function maximisation gives the optimal output quantity vector, y^*, that maximises revenue. The optimal output quantity vector, y^*, uses the fixed input quantity vector x. The revenue function is expressed as a function of exogenous output prices and assumes that they are not zero. The revenue function also depends on the transformation function, which specifies the outputs, y^* that can be produced using the fixed input quantity vector x.

Chambers (1988, p. 263) presents the following properties of the revenue function, $\rho(\boldsymbol{p},\boldsymbol{x})$:

RE1 Non-negativity: $\rho(\boldsymbol{p},\boldsymbol{x}) \geq 0$ for $\boldsymbol{p} > 0$ and $\boldsymbol{x} > 0$;

RE2 Non-decreasing in $\mathbf{p}$: if $\mathbf{p}^1 \geq \mathbf{p}^2$, then $\rho(\boldsymbol{p}^1,\boldsymbol{x}) \geq \rho(\boldsymbol{p}^2,\boldsymbol{x})$;

RE3 Non-decreasing in $\boldsymbol{x}$: If $\boldsymbol{x}^1 \geq \boldsymbol{x}^2$, then $\rho(\boldsymbol{p},\boldsymbol{x}^1) \geq \rho(\boldsymbol{p},\boldsymbol{x}^2)$;

RE4 $\rho(\boldsymbol{p},\boldsymbol{x})$ is convex and continuous in $\boldsymbol{x}$; and

RE5 Positivity and linear homogeneity: $\rho(\varsigma \boldsymbol{p},\boldsymbol{x}) = \varsigma\ \rho(\boldsymbol{p},\boldsymbol{x})$ for $\varsigma > 0$.

These properties parallel those of the profit function.

2.5 Technical, Allocative and Economic Efficiency

As stated in Chapter 1, production is accomplished through one or more processes. In a competitive economic system, the way a firm defines and executes its production processes determines the economic result of the firm, because the quantities of inputs needed for producing outputs depend, not only on the quantities of outputs, but also on the efficiency of the production processes.

In general terms, efficiency is the ability of a process to produce a desired effect, output, cost, or profit, with a minimum of effort, waste, or expense. Farrell (1957) suggested that the efficiency of a firm has two components: technical efficiency and price efficiency. In recent literature, and hereafter, the term, *allocative efficiency,* denotes price efficiency.

Technical efficiency (TE) is associated with how well a firm selects the quantities of inputs to produce given quantities of output, or what outputs it produces from given quantities of inputs, as given by the transformation function. Allocative efficiency (AE) is associated with how well a firm selects the quantities of inputs to produce given quantities of outputs at minimum cost. The combination of technical and allocative efficiencies determines the economic efficiency.

Technical Efficiency

A firm is technically efficient if it produces certain quantities of outputs by using the minimum feasible quantities of inputs or if it produces the maximum possible quantities of outputs for given quantities of inputs.

In terms of inputs, a measure of technical efficiency is defined by a ratio of the minimum feasible quantities of inputs to the actual ones for producing given quantities of outputs. Figure 2.3 represents a technology of one input x to produce one output y. The point P represents an inefficient firm that uses x_1 units of input to produce y_1 units of output. In terms of inputs, the technical efficiency of the firm at P is the ratio, x_2/x_1, where x_2 is the minimum quantity of input required to produce y_1 units of output.

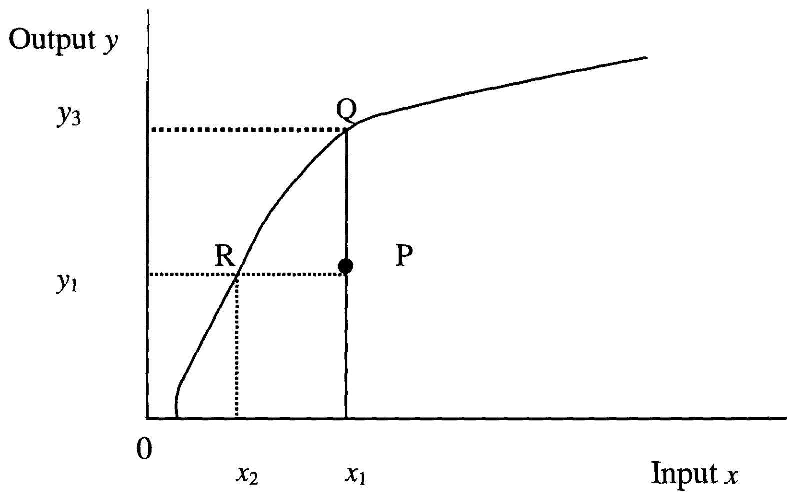

Figure 2.3: Technical Efficiency: Input- and Output-Orientated Measurements

A measure of technical efficiency of a firm, defined in terms of outputs, is defined by a ratio of the actual output quantity to the maximum feasible one for a given input quantity vector. From Figure 2.3, the technical efficiency of the firm at P is the ratio y_1/y_3 where y_3 is the maximum quantity of output that can be produced with x_1 units of input.

In Figure 2.4 below, curve II' represents the unit isoquant that gives the minimum quantities of inputs per unit of output, x_1 and x_2, that fully efficient firms use to produce a unit of output. For the inefficient firm that operates at point P, the technical efficiency is the ratio, OQ/OP, which represents the proportional reduction of inputs required to achieve 100 per cent technical efficiency.

$$TE = OQ/OP. \qquad (2.5.1)$$

Allocative Efficiency

Allocative efficiency (*AE*) requires the firm to select quantities of inputs that produce given quantities of outputs at minimum cost, given observed input prices.[1] In Figure 2.4 the slope of the isocost line, CC', is the input price ratio.[2] For the firm that operates at point P, the allocative efficiency, AE, is the ratio, OR/OQ. The distance RQ represents the reduction in inputs costs, to achieve the cost level associated with the 100 per cent allocative and technical efficient point, Q', i.e.,

$$AE \quad = \quad OR/OQ. \qquad (2.5.2)$$

At point Q', the isocost line CC' is tangent to the isoquant curve II'. This point is technically and allocatively efficient. It is technically efficient because it is on the production function curve, and it is allocatively efficient because the mix of inputs has the minimum technically feasible cost for producing y_0 units of output.

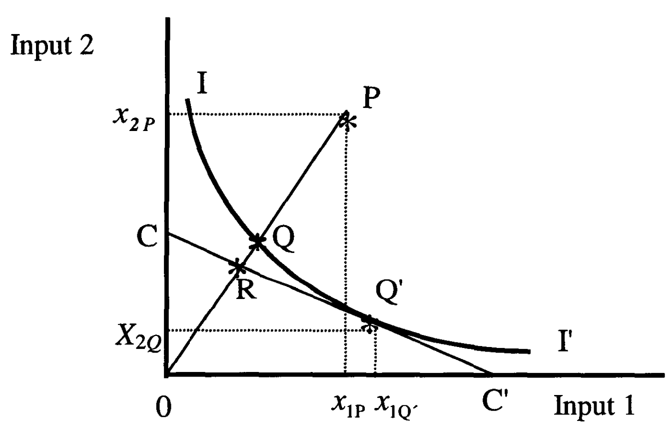

Figure 2.4: Input-Orientated Systems: Technical and Allocative Efficiency where Inputs are Measured per Unit of Output

[1] The efficiency measurements, presented in this thesis, vary between 0 and +1.
[2] For inputs x_1 and x_2 with prices w_1 and w_2, respectively, the isocost equation is: cost = $w_1x_1 + w_2x_2$.

For the multi-input, multi-output case, allocative efficiency is determined by direct evaluation of costs or of revenues, depending on the orientation of the system.

For input-orientated systems, consider a technologically inefficient firm P that uses input quantities, $\boldsymbol{x}_P$, with prices $\boldsymbol{w}$ to produce output quantities $\boldsymbol{y}_P$.[3] The actual cost of inputs is $C_P = \boldsymbol{x}_P\boldsymbol{w}$. By radial reduction of inputs, this firm is technologically efficient using the input quantity vector , $\boldsymbol{x}_Q$, for producing the output quantity vector, $\boldsymbol{y}_P$; at this condition the cost of inputs is $C_Q = \boldsymbol{x}_Q\boldsymbol{w}$. Now let $\boldsymbol{x}_Q'$ be the input quantity vector that produces output quantity vector, $\boldsymbol{y}_P$, with the minimum cost, $C_{Q'} = \boldsymbol{x}_Q'\boldsymbol{w}$. Then, the allocative efficiency of this firm is:

$$AE_I = \frac{C_{Q'}}{C_Q} = \frac{x_{Q'}w}{x_Q w}, \tag{2.5.3}$$

The vector $\boldsymbol{x}_Q'$ satisfies the conditions:

$$\underset{x}{minimise}\ [C_{Q'} = x\ w:\ \ T(y_P, x_{Q'}) = 0], \tag{2.5.4}$$

Then, the solution to (2.5.4) is the system of equations:

$$\begin{cases} w_i = \dfrac{\partial T(\mathbf{y}_P, \mathbf{x})}{\partial x_i}\Big| x_{Q'};\ i = 1,2...I \\ T(\mathbf{y}_P, \mathbf{x}_{Q'}) = 0 \end{cases} \tag{2.5.5}$$

For output-orientated systems, consider a technologically inefficient firm P that uses the input quantity vector, $\boldsymbol{x}_P$, to produce the output quantity vector, $\boldsymbol{y}_P$, that sell at prices $\boldsymbol{p}$. The actual revenue is $R_P = \boldsymbol{y}_P\boldsymbol{p}$. By radial expansion of the outputs, this firm is technologically efficient producing the output quantity vector, $\boldsymbol{y}_Q$, using the input

[3] P, Q and Q', used to describe the input quantity vectors refer to multiple-output equivalents of the input quantity vector shown in Figure 2.4.

quantity vector, x_P; at this condition the revenue is $R_Q = y_Q p$. Now let $y_{Q'}$ be the optimal output quantity vector that maximises revenue, $R_{Q'} = y_{Q'} p$, that input quantities x_P can produce. Then, the output-orientated measurement of allocative efficiency of this firm is:

$$AE_P = \frac{R_Q}{R_{Q'}} = \frac{\mathbf{y}_Q\mathbf{p}}{\mathbf{y}_{Q'}\mathbf{p}}, \tag{2.5.6}$$

The optimal output quantity vector satisfies the condition:

$$\underset{y}{\text{maximise}}[R_{Q'} = yp: \ \mathbf{T}(y_{Q'}, x_P) = 0], \tag{2.5.7}$$

Then, the solution to (2.5.4) is the system of equations:

$$\begin{cases} p_j = \dfrac{\partial T(\mathbf{y}_0, \mathbf{x})}{\partial \mathbf{y}_j} \Big| \ y_{Q'}; \ j = 1,2 \ldots \mathrm{J} \\ T(y_{Q'}, \mathbf{x}_{Q'}) = 0 \end{cases} \tag{2.5.8}$$

Economic Efficiency Measurement

The distance RP, in Figure 2.4, represents the reduction in input costs that the firm at point P must achieve to reduce the cost of inputs to the minimum feasible level (Coelli, Rao and Battese, 1998, p. 135). The economic efficiency (EE) is the ratio of the feasible minimum cost of inputs to the actual one.

$$EE \quad = \quad OR/OP. \tag{2.5.9}$$

From equations (2.5.1) and (2.5.2), it is apparent that:

$$TE \times AE = (OQ/OP) \times (OR/OQ) = OR/OP = EE. \tag{2.5.10}$$

For the multi-input, multi-output case presented for allocative efficiency, the input-orientated economic efficiency measurement is the ratio of the minimum to the actual cost of inputs; then

$$\text{EE}_\text{P}= \frac{\text{C}_{\text{Q}'}}{\text{C}_\text{P}} = \frac{\boldsymbol{x}_{\text{Q}'}\boldsymbol{w}}{\boldsymbol{x}_\text{P}\boldsymbol{w}}, \qquad (2.5.11)$$

and the output-orientated economic efficiency measurement is the ratio of the actual to the maximum revenue; then

$$\text{EE}_\text{P}= \frac{\text{R}_\text{P}}{\text{R}_{\text{Q}'}} = \frac{\boldsymbol{y}_\text{P}\boldsymbol{p}}{\boldsymbol{y}_{\text{Q}'}\boldsymbol{p}} \qquad (2.5.12)$$

To illustrate the concepts developed above, consider the following numerical example. Determine the technical, allocative and economic efficiency of a firm that operates with a production function described by the Cobb-Douglas relationship:

$$y = x_1^{0.6} x_2^{0.4}. \qquad (2.5.13)$$

The relevant data for a firm are:

Output: 115 units,

Input 1: 133 units, price: 2.0 per unit; and

Input 2: 189 units, price: 3.0 per unit.

These data correspond to Firm 4 in the example introduced in Table 3.1 of Section 3.2.

Technical Efficiency Measurement

To measure the technical efficiency of Firm P, we need to know the value of segments $\vec{OP}$ and $\vec{OQ}$, in Figure 2.4. The coordinates of point P are the values of actual inputs x_1 and x_2. The coordinates of point Q define the point where the Cobb-Douglas relationship (2.5.13) and the line OP intersect. The equation of line OP is x_2= (189/133)× x_1. Then, the intersection of line OP and the Cobb-Douglas function is 115 = $[(133/189)\times x_{2Q}]^{0.6}[x_{2Q}]^{0.4}$. From here, x_{2Q}= 141.99, and x_{1Q}= 99.92. [4]

If point P of Figure 2.4 corresponds to the firm, then x_{1P}=133 and x_{2P}= 189.

Then $\vec{OP}=\sqrt{133^2+189^2}$ = 231.11.

Similarly,

$$\vec{OQ}=\sqrt{99.92^2+141.99^2} = 173.62.$$

Finally, the technical efficiency is:

$$TE = 173.62/231.11 = 0.7512.$$

This firm has a technical efficiency of 75.12 per cent. To be 100 per cent efficient, the firm must decrease inputs to:

x_1^*=133×0.7512= 99.91 units and x_2^*= 189×0.7512= 141.98 units.

Allocative Efficiency Measurement

To determine the allocative efficiency of this firm we need to know the value of segments $\vec{OR}$ and $\vec{OQ}$, Figure 2.4. The coordinates of point R are where the line OP intersects with the minimum cost line CC'. As previously indicated, the equation of line OP is $x_1 = (133/189) \times x_2$. The equation of isocost line CC' is:

$$\text{cost} = 2.0 \times x_1 + 3.0 \times x_2.$$

The slope of this line is (-2.0/3.0). The slope of production at the minimum cost of inputs is the same as the slope of the isocost line. This condition gives:

$$-\frac{0.6x_1^{-0.4}x_2^{0.4}}{0.4x_1^{0.6}x_2^{-0.6}} = -\frac{2.0}{3.0};$$

and from this equation $x_1 = 2.25x_2$. After substitution of this relationship in the Cobb-Douglas production function, we obtain, $x_{1Q'} = 159.06$ units and $x_{2Q'} = 70.69$ units. The equation of the minimum cost isocost line is $530.21 = 2.0 \times x_1 + 3.0 \times x_2$. As stated above, the intersection of this equation and line OP gives the coordinates of point R. Then $x_{1R} = 90.76$ and $x_{2R} = 128.97$.[5]

Also

$$\vec{OR} = \sqrt{90.76^2 + 128.97^2} = 157.70.$$

Finally, the allocative efficiency is:

$$\text{AE} = 157.70/173.63 = 0.9082.$$

This firm has an allocative efficiency of 90.82 per cent.

Economic Efficiency Measurement

To determine the economic efficiency of this firm, we need to know the value of segments $\vec{OP}$ and $\vec{OR}$ in Figure 2.4. From the technical and allocative efficiency calculations, we know that the value of these segments are 231.11 and 157.70 respectively. Finally, the economic efficiency is:

---

[4] The coordinates x_{1Q} and x_{2Q} are not shown in Figure 2.4.

$EE = 157.70/231.11 = 0.6824$

This firm has an economic efficiency of 68.24 percent ($= 0.7512 \times 0.9082$).

2.6 Productivity

Kendrick (1993) states that: "Productivity is the ratio of output to inputs of labour and other resources, in real terms." This is the concept that is most frequently used by authors (Thor, 1993, pp. 2-9.1; Lovell, 1993, p. 3; Førsund, 1993, p. 352).

Grosskopf (1993, p.162) states an operational definition of productivity: "By total factor productivity I mean an index of output divided by an index of total input usage. Thus total factor productivity is a generalisation of single-factor productivity measures, such as labor productivity which is the ratio of (an index of) output to a single input, labour." Two concepts deserve special attention in this definition.

The first concept is that *productivity* is the ratio of aggregated outputs to aggregated inputs, instead of the ratio of just outputs and inputs. There are different options to formulate the aggregation. For a single-input, single-output process, productivity may be measured as the ratio of the physical quantities. For the purposes of this thesis, for a multiple-input, multiple-output system the aggregation of inputs is a weighted summation of inputs and the same for outputs. The addition of weighted inputs must be done in some system of units; for this reason the units of the weight factors include the reciprocal of the units of the corresponding physical quantities of inputs and outputs.

---

[5] The coordinates x_{1R} and x_{2R} are not shown in Figure 2.4.

Prices are possible weight factors. In this way, physical quantities of inputs (and outputs) are changed to a one-dimensional system. This form of aggregation of inputs and outputs is used in the DEA presentation in Chapter 3 of this thesis.

The second concept is that the measurement of productivity may be done with respect to total input usage, *total factor productivity* (TFP), or with respect to some particular input, *partial productivity*.

Under weighted summation for aggregation of inputs and outputs, the total factor productivity of firm e is:

$$\text{TFP}_\text{e} = \frac{\boldsymbol{u}_e \boldsymbol{y}_e}{\boldsymbol{v}_e \boldsymbol{x}_e}, \qquad (2.6.1)$$

where for firm e, $\boldsymbol{u}_e$ is the weight vector of outputs; $\boldsymbol{v}_e$ is the weight vector of inputs; $\boldsymbol{x}_e$ is the input quantity vector; and $\boldsymbol{y}_e$ is the output quantity vector.

The main problem with the definition of equation (2.6.1) is that weights $\boldsymbol{u}_e$ and $\boldsymbol{v}_e$ are unknown. The value of these weights must be defined by management, or determined using some methodology. Frequently, output and input prices are used as weight factors. Chapter 3 presents DEA as a non-parametric method for determining the values of these weights, when measuring the relative productivity of a firm.

For a production process, the partial productivity is the ratio of a weighted summation of all outputs to some specific input (or a set of specific inputs). Then, the partial productivity (PP) of firm e with respect to input i is:

$$PP_{ie} = \frac{u_e y_e}{x_i}, \qquad (2.6.2)$$

The terms *productivity* and *total factor* productivity are used interchangeably in this thesis.

2.7 Dynamic Models

In previous sections, the main assumption is that the systems involved are static, in the sense that the state of technology is constant over time. This means that although inputs and outputs may change over time, the technology is the same before and after the adjustments are done. The transformation function represents the technology. This section presents basic concepts of *technical change*. Chambers (1988, p. 205) defines technical change as the shift in the production function over time, as a stable relationship between output, inputs and time. Technical change is measured by how output changes as time elapses, with the input quantity vector held constant.

For the purpose of this section, technology innovation and new technology are different. As example, consider mechanical typewriters and word processor computers and printers. The last are new technology, not just an innovation. The technical innovation is *embodied* in this new technology, in the sense that a new sort of machines is required to have access to the new technology. In this thesis, we consider *disembodied technical change* only.

Dynamics of Efficiency I

This subsection presents basic concepts of *dynamics of efficiency*, with special focus on the expansion frontier and allocative efficiency. The presentation is mainly based on Sengupta (1995, Section 2.1).

For the one-output and multiple-inputs case, as stated above, the Cobb-Douglas form is most often used for the production function. Equation (2.2.2a) is a static production function because it does not include time.

Restating the cost-minimisation problem of equation (2.4.2), with the Cobb-Douglas equation (2.2.2a) replacing the general transformation function, the cost-minimisation problem takes the form:

$$c(\boldsymbol{w},\boldsymbol{y}) = \underset{\boldsymbol{x}}{\text{minimise}}\{\mathbf{w}\boldsymbol{x}:\ y^* = \alpha_0 \prod_{i=1}^{I} x_i^{\alpha_i}\}. \tag{2.7.1}$$

In terms of input one, x_1, the solution to problem defined by (2.7.1) is:

$$\mathbf{x}_\mathrm{i}^* = \frac{\alpha_\mathrm{i}\omega_I}{\alpha_I\omega_\mathrm{i}}\mathbf{x}_1^*\ ,\ i=1, 2, 3,..., \mathrm{I}; \tag{2.7.2}$$

where x_i^* is the quantity of input i that minimises the total cost of inputs; ω_i is the price of input i; and α_i is the exponent of input i in the Cobb-Douglas function.

Setting $r_{1i} = \dfrac{\alpha_i\omega_1}{\alpha_1\omega_i}$, equation (2.7.2) becomes

$$x_i^* = r_{1i}\ x_1^*. \tag{2.7.3}$$

Replacing x_i^* in the Cobb-Douglas equation (2.2.2a), the mathematical formulation for the optimal output y^*, is given below in equation (2.7.4).

$$y^* = \alpha_0\ \boldsymbol{x}_1^{*\alpha_I} \prod_{i=2}^{I} r_{1i}\ x_1^{*\alpha_i} \tag{2.7.4}$$

The optimal output y^* on this expansion frontier follows the trajectory as long as prices and the α_i-parameters are constant over time (Sengupta, 1995, p. 39). When outputs and inputs are time-varying, a dynamic case is obtained, and the factor, r_{1i}, is called the expansion frontier ratio (Sengupta, 1995, p. 39). In a time-derivative form, the output trajectory becomes equation (2.7.5)

$$\frac{\dot{y}_t^*}{y_t^*} = \frac{\dot{\alpha}_0}{\alpha_0} + \alpha_1 \frac{\dot{x}_1^*}{x_1^*} + \sum_{i=2}^{I} \frac{\alpha_i}{r_{1i} x_1} (x_1 \dot{r}_{1i} + r_{1i} \dot{x}_1), \tag{2.7.5}$$

where the dot denotes the time derivative. The expansion frontier ratio, $r_{1i} = x_1^* / x_i^*$, obtained from equation (2.7.3), and the optimal output trajectory, given by equation (2.7.5), have two main implications (Sengupta, 1995, pp. 40-41).

The first implication of equation (2.7.5) is that, if input prices are constant, the ratio of input quantities, r_{1i}, is constant. In addition, if α_0 is constant, then the frontier is given by:

$$\frac{\dot{y}_t^*}{y_t^*} = \frac{\dot{x}_I^*}{x_I^*} \sum_{i=1}^{I} \alpha_i . \tag{2.7.6}$$

Under these conditions, the boundary of technology is static and the time rate of change of output, per unit of output quantity, increases or decreases in proportion to the time rate of change of quantity per unit of x_1.[6] The proportionality factor equals $\sum_{i=1}^{I} \alpha_i$. For Cobb-Douglas technology, this proportionality factor gives the returns to scale of production function. If $\sum_{i=1}^{I} \alpha_i$ is larger than unity, the technology exhibits increasing returns to scale; if $\sum_{i=1}^{I} \alpha_i$ is equal to unity, the technology exhibits constant returns to scale; and if $\sum_{i=1}^{I}\alpha_i$ is less than unity, the technology exhibits decreasing returns to scale.

In economic growth models, this case is analysed in terms of the von Neumann optimal trajectory (Sengupta, 1995, p. 40).

The second implication of equation (2.7.5) is that it shows how the efficient firms respond to the time rate of change in α_0, input prices, and x_1^*. In this case, the boundary of technology shifts. For the case of two inputs, a log-log plot of the production function shows a family of parallel lines (constant α_1 and α_2), shifting with the intercept, log(α_0). This case is called the Hicks-neutral technical progress (Sengupta, 1995, p. 41).

Dynamics of Efficiency II

Färe and Grosskopf (1996, pp. 152-153) present another approach to dynamic technical boundary. Consider the general transformation function, defined by equation (2.3.1). For the case of a series of observations, dated at t = 1, 2, 3,...,T, the transformation function may be written at time t as follows:

$$T_t(\boldsymbol{x}_t,\ \boldsymbol{y}_t^*) = 0. \qquad (2.7.7)$$

A pair,($\boldsymbol{x}_t, \boldsymbol{y}_t^*$), consist of all input and output quantity vectors such that the input quantity vector $\boldsymbol{x}_t$ can produce the output quantity vector $\boldsymbol{y}_t^*$. For this reason, $T_t(\boldsymbol{x}_t, \boldsymbol{y}_t^*) = 0$ may be regarded as a series of static transformation functions. This

[6] Recall that input one is selected to serve as the referent of changes over time.

consideration is fundamental for the formulation of the basic DEA model for the selection of optimal paths of adjustment to be presented in Chapter 4.

Conclusion

From this section, following Chambers (1988), Sengupta (1995) and Färe and Grosskopf (1996), we conclude that there are models for dynamic transformation functions, which define dynamic boundaries of technology. These models are used mainly for explaining or for comparing the past performance of firms. Nonetheless, we use dynamic transformation functions in Chapter 4 for forecasting the behaviour of firms in a short time horizon.

2.8 Primal and Dual Dynamic Models

This section presents some basic primal and dual dynamic models that extend the static theory of the firm. As indicated by Wayne and Shumway (1988, p. 837), dynamic models that are consistent with the theory of the firm have been derived from applications of optimal control theory, and have been applied in various forms, primarily in partial adjustment of inputs and outputs. Primal and dual models are derived from the inter-temporal value function, which is the present value of a stream of future profits (or of costs). The behavioural equations may be obtained via the primal approach by using the first-order Euler equation, or by application of the envelope theorem to the value function.

Primal Models

The primal approach was presented first by Treadway (1970). Berndt, Fuss and Waverman (1981) applied this approach to U. S. manufacturing industry. Wayne and Shumway (1988, p. 837) indicate that primal models are limited to modelling one quasi-fixed input, or assuming independent adjustment between two or more quasi-fixed inputs.

The theory of the profit-maximising firm has been established on reasonably and logical bases, considering static or atemporal optimisation concepts. Nonetheless, this theory has served as the foundation for intensive econometric work on input-demand and product-supply. The atemporal constraint makes it necessary to incorporate specific adjustment mechanisms. Some attempts to provide a consistent optimal dynamic theory of the firm have been constrained to reproduce the inferences of the static theory, even in those cases that dynamic concepts generate predictions Treadway (1970, p. 329).

Treadway (1970) formulates and analyses a general dynamic model seeking new principles, even questioning some theorems of static production theory. As an example, he shows that in a dynamic model involving adjustment cost, under some circumstances, the static profit-maximisation rule of decreasing long-run product-supply and increasing long-run input-demand curves is not valid. Treadway (1970, p. 330) indicates that (sic) "there is no necessity for a static production function constraint in a dynamic optimization theory."

The dynamic model of Treadway (1970, p. 331) is to maximise the (functional) present value of the competitive firm, which depends on time paths of product, $y(t)$, labour services, $\boldsymbol{x}(t)$, and real investment, $I(t)$. The corresponding constant prices are $\boldsymbol{p}$, $\boldsymbol{w}$, and g. Then,

$$V[\boldsymbol{y}(t), \boldsymbol{x}(t), \boldsymbol{I}(t)] = \int_0^\infty [\boldsymbol{py}(t) - \boldsymbol{wx}(t) - gI(t)]e^{-rt}\, dt, \tag{2.8.1}$$

where r is the rate of discount for the value of money, and t is time.

Assuming that depreciation is proportional to the fixed assets, K, and that the fixed assets at t=0 have value, $K(0)=K_0$, an economic constraint to (2.8.1) is the net rate of investment, $\dot{K}$, defined by

$$\dot{K} = I(t) - \mu K(t) \qquad \mu \geq 0. \tag{2.8.2}$$

Finally, Treadway (1970) assumes that the production function is such that the output quantity vector, $\boldsymbol{y}$, depends not only on the fixed asset, K, and of the input quantity vector, $\boldsymbol{x}$, but also on $\dot{K}$,

$$\boldsymbol{y} = f(K, \boldsymbol{x}, \dot{K}). \tag{2.8.3}$$

Treadway (1970, p. 332) justifies the incorporation of the net rate of investment in the production function arguing that it is embodied in (sic) "the assumption that the magnitude of capital stock can be changed only by incurring adjustment costs."

Treadway (1970, p. 345) points out the following main conclusions derived from his intertemporal optimisation model: first, the demand of investment, $I(t)$, is inversely related to the fixed assets, $K(t)$, and to the real rental, $(r+\mu)g$, in a neighbourhood of a long-run equilibrium point. This means that the model gives a theory of investment behaviour that is essentially dynamic. Second, it is not possible (sic) "to demonstrate (a) the non-positivity of own-wage effects on the demand for the variable factor either in the short-run or the long-run; (b) the non-negativity of supply effects of a change in product price in either the short-run or the long-run; or (c) that short-run effects are less elastic than long-run effects".

Dual Models

The dual approach is derived from the results presented first by McLaren and Cooper (1980), and Taylor and Cooper (1980), and formalised by Epstein (1981). Other applications of this model are: Epstein and Denny (1983) who use this approach for empirical studies of U.S. manufacturing industry, Taylor (1984), Taylor and Monson (1985), Vasavada and Chambers (1986), Luh and Stefanou (1991), and Luh and Stefanou (1996) who apply the model to U.S. agriculture.

Wayne and Shumway (1988) use a dual model for modelling more than one quasi-fixed input and for testing for independent adjustment, rather than simply assuming it, as it is done in the primal model.

For the purpose of this thesis, instead of reviewing the paper that initiated the study of dynamic dual models, we review the main concepts presented by Luh and Stefanou (1996), who study the decisions behaviour of competitive firms, based on their perception of future market opportunities, to maximise the present value of profits over an infinite time horizon. The firms face internal adjustment costs. Consider the present value function dynamic model:

$$\mathrm{V}[\boldsymbol{p}, \boldsymbol{w}, \boldsymbol{g}, \boldsymbol{k}, t] = \int_{t}^{\infty} [\boldsymbol{p}(\tau)\boldsymbol{y}(\tau) - \boldsymbol{w}(\tau)\boldsymbol{x}(\tau) - \boldsymbol{g}(\tau)\boldsymbol{K}(\tau)] e^{-r(\tau-t)} \mathrm{dt}, \qquad (2.8.4)$$

where $p(\tau)$, $w(\tau)$, and $g(\tau)$ denote, respectively, prices of output quantity vector, $\boldsymbol{y}(\tau)$, prices of the variable input quantity vector, $\boldsymbol{x}(\tau)$, and prices of vector of quasi-fixed inputs, $\boldsymbol{K}(\tau)$, at time τ. The constant real discount rate of worth of money is r. Equation (2.8.1) has three decision variables: $\boldsymbol{y}(\tau)$, $\boldsymbol{x}(\tau)$, and the vector of gross investment, $\boldsymbol{I}(\tau)$, associated with the quasi-fixed inputs, $\boldsymbol{K}(\tau)$; one endogenous state variable, $\boldsymbol{K}(\tau)$, and three exogenous state variables, $\boldsymbol{p}(\tau), \boldsymbol{w}(\tau)$, and $\boldsymbol{g}(\tau)$.

The firm's expected time evolution of the state variables is described by a set of first-order ordinary differential equations:

$$\dot{\boldsymbol{K}}(\tau) = \boldsymbol{I}(\tau) - \delta \boldsymbol{K}(\tau) ; \boldsymbol{k}(t)= \boldsymbol{k} \qquad (2.8.5)$$

$$\dot{p}(\tau) = \alpha + \theta_\alpha p(\tau) ; p(t)= p \qquad (2.8.6)$$

$$\dot{w}(\tau) = \beta + \theta_\beta w(\tau) ; w(t)= w \qquad (2.8.7)$$

$$\dot{g}(\tau) = \gamma + \theta_\gamma g(\tau) ; g(t)= g, \qquad (2.8.8)$$

where δ is the vector of depreciation rates of fixed assets $\boldsymbol{K}$. The ordinary differential equations (2.8.6), (2.8.7) and (2.8.8) represent the expected evolution of prices. For the case of stationary prices, these equations are set to value zero and functions α, β, and γ do not include time.

Non-static technology is considered by making explicit the dependence of the production function on time:

$$\mathrm{y}(\tau)= \mathrm{f}[\, \boldsymbol{x}(\tau), \boldsymbol{k}(\tau), \ \boldsymbol{I}(\tau), \tau], \ \forall \tau \in [t, \infty) \qquad (2.8.9)$$

Luh and Stefanou (1996, pp. 996-997) solve the differential equations (2.8.6), (2.8.7) and (2.8.8) and the solution is presented in the form of discrete-time transition equations:

$$p(\tau) = \bar{\alpha} + \bar{\theta}_{\alpha}\, p(\tau-1) \tag{2.8.10}$$

$$w_i(\tau) = \bar{\beta} + \bar{\theta}_{\beta,}\, w_i(\tau-1) \tag{2.8.11}$$

$$g_i(\tau) = \bar{\gamma} + \bar{\theta}_{\gamma,}\, g_i(\tau-1) \tag{2.8.12}$$

Equations (2.8.10), (2.8.11) and (2.812) represent the expected variation of prices, from one period to the next.

The duality between the value function (2.8.4) and the production function (2.8.9) permits the authors to derive a system of equations that rule the firms´ estimation of output supply and inputs demand under non static price expectations.

The methodology developed by the authors incorporates non-static expectations into dynamic analysis. Nonetheless, the theoretical model assumes that the competitive firm determines rationally how future prices evolve, and that prices are known with perfect certainty at each decision time.

2.9 Conclusions

Technical efficiency can be measured when quantities of inputs and outputs are the only available data. Allocative efficiency can be measured when quantities and prices of inputs and outputs are available. Economic efficiency is of special interest for the purposes of this thesis. To the extent that most retailing stores minimise input cost, for a given output, the input quantities must be adjusted to operate at a multiple-input, multiple-output equivalent to point Q' of Figure 2.4. Technical efficiency and total factor productivity are two basic measurements of the performance of a productive process.

For the purpose of this thesis, the optimisation of profit, cost, and revenue functions is fundamental for developing the objective function of the optimal path of adjustment models.

Finally, there are primal and dual models for dynamic production functions that define dynamic boundaries of technology and present value functions that maximise the profits of firms. The dynamic production functions are used for explaining or for comparing the past performances of firms and for predicting dynamic demand supply schedules.