# CHAPTER 3

# FUNDAMENTALS OF DATA ENVELOPMENT ANALYSIS

## 3.1 Introduction

Chapter 2 presents basic concepts of production and technical efficiency. It discusses the production and transformation functions and considers profit maximisation, cost minimisation and revenue maximisation as possible behavioural motivations for producers. Finally, it introduces the concepts of technical, allocative, and economic efficiency.

The main assumption underlying the formulation presented in Chapter 2 is that the production (or transformation) functions are known. Production and transformation functions may be estimated using parametric methods, of which ordinary least squares regression is the most commonly used. The main problem with this approach is that the selected functional form may be misspecified. Non-parametric methods can be used to overcome this problem (Thanassoulis, 2001, p. 9).

In this chapter, Data Envelopment Analysis (DEA) is introduced as a benchmarking, non-parametric method, which can be used to optimise objective functions of special interest and obtain boundary solutions that define 100 per cent technical efficiency.

Also, DEA is introduced as a method that is well suited for determining the optimal quantities of inputs (or outputs) that optimise some objective function of interest such as revenue, cost, and present value of cash flows. This method is equally valid for firms with multiple inputs and multiple outputs, and does not require the specification of a functional relationship between inputs and outputs.

In general terms, DEA is a non-parametric method for measuring relative efficiency of members of a set of comparable firms. Because the production function of efficient firms is not known *a priori*, it must be estimated using observed sample data from the industry involved. For two inputs and one output, the best-performing firms define piecewise linear boundaries. For multiple outputs and multiple inputs, the best-performing firms define piecewise surfaces that are hyperplane boundaries (Coelli, Rao and Battese, 1998, p. 140).

Section 3.2 states the primal DEA problem and discusses productivity that is defined as the largest value of a weighted sum of outputs divided by a weighted sum of inputs. The weights are obtained using linear programming. A simple example illustrates these concepts. The section 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 returns to scale concept. Section 3.5 reviews, in DEA formulation, the expected optimising behaviour of firms. Section 3.6 discusses some limitations of DEA. Finally, Section 3.7 presents some concluding comments.

## 3.2 Productivity and the Primal DEA Problem

This section presents basic concepts associated with DEA, followed by the DEA primal formulation that is related to the productivity concept. Concepts and models are illustrated with a simple example.

DEA is a non-parametric method for extracting information from a collection of observed production data on a set of comparable firms. DEA is non-parametric, because no functional relationship is assumed between outputs and inputs. DEA optimises the performance of each individual firm, relative to all other firms in the set of selected firms. Productivity, which is the ratio of a weighted sum of outputs to the weighted sum of inputs, is the measure to be maximised in this section.

Charnes, Cooper and Rhodes (1978) extended Farrell's ideas, linking the estimation of technical efficiency with production frontiers, in a non-parametric approach. Their model, known as the CCR model, generalised a "single-output to a single-input" ratio measure of productivity for a single firm to a "multiple-output to multiple-input" ratio measure of productivity, relative to a set of firms.

Later, Charnes and Cooper (1985) state their concept of relative efficiency, signalling that DEA is the method for relative productivity measurement. The productivity measurements that are presented in Section 2.6 and in this chapter are different, although related. The first one is a performance measure of a firm, isolated from other firms; the second is a performance measure of a firm relative to the performance of a set of comparable firms.

*The Primal DEA Model*

To formulate the basic mathematical primal problem, consider a simple set of five firms ($k$ = 1, 2, 3, 4, 5). Each firm produces quantities of one output, $y_1, y_2, y_3, y_4, y_5$, using quantities of two inputs, represented by the vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5$. In scalar form, we define the productivity of firm $e$ (equation 2.6.1) as: [1]

$$\textit{Total Factor Productivity}_e = \frac{u_1 y_e}{v_1 x_{1e} + v_2 x_{2e}}. \qquad (3.2.1)$$

The main interest is to determine the values of the weights, $\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{v}_2$, that maximise the *Total Factor Productivity*$_e$ measure in equation (3.2.1). The optimisation is subject to the constraint that no firm may have *Total Factor Productivity* larger than 1.0, when its productivity is evaluated using the weights that optimise the productivity measurement for firm $e$. The optimisation problem becomes:

$$\text{maximise } \frac{u_1 y_e}{v_1 x_{1e} + v_2 x_{2e}} \qquad (3.2.2)$$

subject to

$$\frac{u_1 y_k}{v_1 x_{1k} + v_2 x_{2k}} \leq 1.0, \text{ for } k = 1, 2, 3, 4, 5,$$

$$u_1, v_1, v_2 \geq 0.$$

[1] Firm $e$ is a specific firm under evaluation. Firm $e$ is a member of the set of firms under study.

The solution to (3.2.2) is not unique, in the sense that if $(u_1^*, v_1^*, v_2^*)$ is an optimal solution, then $\alpha(u_1^*, v_1^*, v_2^*)$ is another solution. To overcome this, the objective function is split into the numerator and the denominator, and the numerator is maximised, while the denominator is forced to have the value one.[2] The optimisation problem then becomes:

$$\text{maximise } u_1 y_e \qquad (3.2.3)$$

subject to

$$v_1 x_{1e} + v_2 x_{2e} = 1$$

$$u_1 y_{1k} - (v_{1k} x_{1k} + v_{2k} x_{2k}) \leq 0 \text{ for } k = 1, 2, 3, 4, 5,$$

$$u_1, v_1, v_2 \geq 0.$$

Problem (3.2.3) is a linear programming problem.[3]

In problem (3.2.3) the subscript, $e$, stands for the firm under evaluation, $e = 1, 2, \ldots, 5$. For this reason, to evaluate the productivity of the five firms requires the successive solution of problem (3.2.3) for each of the five firms.

To illustrate the concepts presented above, consider the data of Table 3.1 and for each firm find the output productivity, as defined with problem (3.2.3).

---

[2] This produces what many researchers call "output-orientated" performance measures. To obtain input-orientated measures the denominator is minimised and the numerator is constrained to equal one. The issue of orientation is discussed below.

[3] We note that this LP implicitly assumes that the production technology has constant returns to scale. The issue of returns to scale and scale efficiency are discussed below.

**Table 3.1:** Input and Output Data for Five Firms

| Firm $k$ | Quantity $y_k$ | $x_{1k}$ | $x_{2k}$ |
|---|---|---|---|
| 1 | 100 | 100 | 100 |
| 2 | 110 | 90 | 149 |
| 3 | 120 | 150 | 85.9 |
| 4 | 115 | 133 | 189 |
| 5 | 103 | 152 | 61 |

Using (3.2.3), the LP problems for firms 1 and 4, are:

| Firm 1: | Firm 4: |
|---|---|
| maximise $100u_1$ | maximise $115u_1$ |
| subject to: | subject to: |
| $100v_1+100v_2=1$ | $133v_1+189v_2=1$ |
| $100u_1 - 100v_1 - 100v_2 \leq 0$ | $100u_1 - 100v_1 - 100v_2 \leq 0$ |
| $110u_1 - 90v_1 - 149v_2 \leq 0$ | $110u_1 - 90v_1 - 149v_2 \leq 0$ |
| $120u_1 - 150v_1 - 85.9v_2 \leq 0$ | $120u_1 - 150v_1 - 85.9v_2 \leq 0$ |
| $115u_1 - 133v_1 - 189v_2 \leq 0$ | $115u_1 - 133v_1 - 189v_2 \leq 0$ |
| $103u_1 - 152v_1 - 61v_2 \leq 0$ | $103u_1 - 152v_1 - 61v_2 \leq 0$ |
| $u_1, v_1, v_2 \geq 0$ | $u_1, v_1, v_2 \geq 0$ |

Similar LP problems are required for firms 2, 3 and 5. Table 3.2 shows the solution to problem (3.2.3) for the five firms.

**Table 3.2:** Solution to Problem (3.2.3) for the Five Firms:

Output-Orientated System and Constant Returns-to-Scale Technology

| Firm | Productivity | $100u_1$ | $1000v_1$ | $1000v_2$ |
|---|---|---|---|---|
| 1 | 1.0 | 1.0 | 6.61 | 3.39 |
| 2 | 1.0 | 0.909 | 6.01 | 3.08 |
| 3 | 1.0 | 0.833 | 4.43 | 3.90 |
| 4 | 0.757 | 0.658 | 4.35 | 2.23 |
| 5 | 1.0 | 0.9709 | 0 | 16.39 |

Following Thanassoulis (2001, p. 93), the information obtained from this DEA model is:

Firms 1, 2, 3 and 5 are fully efficient, having maximum productivity of 1.0.

Firm 4 is not efficient, having a relative productivity of 0.757, which implies that it is 75.7 per cent efficient. To become fully efficient, Firm 4 must increase its output from the actual quantity of 115 units to 115/0.757 = 151.8 units. The output of 151.8 units is the largest quantity technically attainable, using 133 and 189 units of inputs 1 and 2, respectively.

*Input- and Output-Orientated Measures*

It should be noted that the above performance measures are output-orientated measures, i.e., they look at the amount by which observed output falls short of potential output, for a fixed vector of inputs. One can also calculate input-orientated measures, which look at the amount by which observed inputs exceed the minimum possible levels, for a fixed vector of outputs.

Input-orientated measures are obtained by taking problem (3.2.2) and, instead of maximising the weighted sum of outputs (as for the output-orientated case), we minimise the weighted sum of inputs. The resulting input-orientated LP is:

$$\text{minimise } \boldsymbol{v}_e \boldsymbol{x}_e \qquad (3.2.4)$$

subject to

$$\boldsymbol{u}_e \boldsymbol{y}_e = 1$$

$$\boldsymbol{u}_e \boldsymbol{y}_k - \boldsymbol{v}_e \boldsymbol{x}_k \leq 0 \qquad k = 1, 2, ..., \text{K},$$

$$\boldsymbol{u}_e, \ \boldsymbol{v}_e \geq 0.$$

Note that the productivity measures derived from this LP will be identical to the ones obtained from the output-orientated LP presented in problem (3.2.3). Thus, it appears that the orientation issue is purely an academic one. This is true in the case of constant returns-to-scale technology, which we consider in this section. However, when we look at alternative DEA models, such as the variable returns-to-scale model, the issue becomes important because the chosen orientation can affect the measures obtained.

Given this information, how can one select the correct orientation? In most studies, the selection of orientation tends to depend on an assessment of which variables the managers have most control over. For example, in firms where we observe that the firms have set orders to fill, the input quantities appear to the primary decision variables. Examples are water supply firms and electric power generation. However, in some other industries, firms may have fixed quantities of resources and be asked to produce as much output as possible. Examples are farms with fixed land areas, and small and medium sized stores with fixed staff numbers.

### 3.3 Technical Efficiency and the DEA Dual Problem

This section presents the DEA dual formulation of problems (3.2.3) and (3.2.4), and relates them to technical efficiency concepts.[4] Concepts and models are illustrated with simple examples. Using the duality property of linear programming, the dual formulation of problems (3.2.3) and (3.2.4) are directly derived. Nonetheless because productivity and technical efficiency are two strong concepts that are frequently used in production economics, the DEA dual formulation of those two problems is derived in detail.

As stated in Section 2.5, "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. In terms of outputs, a measure of technical efficiency of a firm is the ratio of the actual output quantities to the maximum feasible ones, where the actual and maximum feasible output quantities are for the same quantities of inputs.

*Input-Orientated Technical Efficiency Measurement*

To formulate those concepts in mathematical form, consider a set of K firms of an input-orientated system with constant returns-to-scale technology. Each firm produces

[4] A mathematical presentation and analysis of the relationships between the primal and dual formulations of an LP is in Chapter 8 of Hadley (1972).

J outputs, at quantities, $y_k$, $k$=1, 2, …, K, respectively, using $I$ inputs having values, $\boldsymbol{x}_k$, $k$=1, 2, …, K.

Assuming that there are no slacks involved, using equation (2.5.1) for input-orientation measurement of the technical efficiency of firm $e$, the optimal input quantities, $\boldsymbol{x}_e^*$, may be expressed as the product of the technical efficiency measurement, $TE_{ie}$, times the actual inputs, $\boldsymbol{x}_e$, as follows:

$$\boldsymbol{x}_e^* = TE_{ie}\ \boldsymbol{x}_e . \qquad (3.3.1)$$

The optimal input quantities, $\boldsymbol{x}_e^*$, are unknown. DEA specifies $\boldsymbol{x}_e^*$ as a sum of weighted values of optimal quantities, $\boldsymbol{x}_k^*$, $k$ = 1, 2, …, K. Under these conditions, with $\lambda_k$ being the weight of firm $k$, the right-hand side of equation (3.3.2) is the minimum feasible vector of input quantities, and the right-hand side of equation (3.3.3) is the maximum feasible vector of output quantities:

$$\boldsymbol{x}_e^* = TE_{ie}\ \boldsymbol{x}_e \geq \boldsymbol{X}\boldsymbol{\lambda} \qquad (3.3.2)$$

$$\boldsymbol{y}_e \leq \boldsymbol{Y}\boldsymbol{\lambda} , \qquad (3.3.3)$$

where $\boldsymbol{X}$ is the I×K matrix of observed input quantities, and $\boldsymbol{Y}$ is the J×K matrix of observed output quantities. The vectors, $\boldsymbol{x}_e$ and $\boldsymbol{y}_e$ are the known input-output data of firm $e$, which is one of the K firms. Together, inequalities (3.3.2) and (3.3.3) are the DEA expression for the boundary of technology.

Mathematically, in order for $\boldsymbol{x}_e^*$ to be a minimum, $TE_{ie}$ has to be a minimum. The LP to minimise $TE_{ie}$ is presented in problem (3.3.4):

$$\text{minimise } TE_{ie} \qquad (3.3.4)$$

subject to:

$$X\lambda \leq TE_{ie}\ x_e$$

$$Y\lambda \geq y_e$$

$$\lambda \geq 0,$$

where $TE_{ie}$ is the input-orientated technical efficiency measurement for firm *e*. The main interest is to determine the values of the weights, $\lambda$, that minimise $TE_{ie}$. Equations (3.3.2) and (3.3.3) set constraints to the decision vector $\lambda$.[5]

In problem (3.3.4) the subscript, *e*, stands for the firm under evaluation, $e = 1, 2, \ldots, K$. For this reason, to evaluate the productivity of the K firms requires the successive solution of problem (3.3.4) for each of the K firms. This model was first developed in Charnes, Cooper and Rhodes (1978).

Consider the data of Table 3.1 and for each firm find the input-orientated technical efficiency, $TE_{ie}$, as defined with problem (3.3.4). The LP problems for firms 1 and 4, input orientated systems, and constant returns-to-scale technology are:

| Firm 1: | Firm 4: |
|---|---|
| minimise $TE_{i1}$ | minimise $TE_{i4}$ |
| subject to: | subject to: |
| $100\lambda_1+110\lambda_2+120\lambda_3+115\lambda_4+103\lambda_5 \geq 100$ | $100\lambda_1+110\lambda_2+120\lambda_3+115\lambda_4+103\lambda_5 \geq 115$ |
| $100\lambda_1+90\lambda_2+150\lambda_3+133\lambda_4+152\lambda_5 \leq 100TE_{i1}$ | $100\lambda_1+90\lambda_2+150\lambda_3+133\lambda_4+152\lambda_5 \leq 133TE_{i4}$ |
| $100\lambda_1+149\lambda_2+85.9\lambda_3+189\lambda_4+61\lambda_5 \leq 100TE_{i1}$ | $100\lambda_1+149\lambda_2+85.9\lambda_3+189\lambda_4+61\lambda_5 \leq 189TE_{i4}$ |
| $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0$ | $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0$ |

[5] In equations (3.3.2) and (3.3.3), $y_e$, $Y$ and $X$ contain the observed data.

Similar LP problems are required for firms 2, 3 and 5. Table 3.3 shows the solution for the five firms. In Section 2.5, and assuming that the production function for this industry has a Cobb-Douglas form, the technical efficiency measurement for firm 4 is 0.751.[6] This measurement differs by less than 0.8 per cent with the technical efficiency measured with DEA. Note that the productivity measurement for firm 4 and the technical efficiency measurement for the same firm is exactly the same value, 0.757. At the optimum, the value of the objective function of the primal LP and the value of the objective function of the corresponding dual LP is the same.

**Table 3.3:** Solution to Problem (3.3.4) for the Five Firms: Input-Orientated Technical Efficiency Measurement and Constant Returns-to-Scale Technology

| Firm | Technical Efficiency | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|
| 1 | 1.0 | 1.0 | 0 | 0 | 0 | 0 |
| 2 | 1.0 | 0 | 1.0 | 0 | 0 | 0 |
| 3 | 1.0 | 0 | 0 | 1.0 | 0 | 0 |
| 4 | 0.757 | 0.360 | 0.718 | 0 | 0 | 0 |
| 5 | 1.0 | 0 | 0 | 0 | 0 | 1.0 |

[6] Note that data for the example of Section 2.5 correspond to data of firm 4 in Table 3.1

Again, following Thanassoulis (2001, p. 93), the information obtained from this DEA model is:

Firms 1, 2, 3 and 5 are fully efficient, having maximum technical efficiency. Firm 4 is not efficient, having a relative efficiency of 0.757. This is equivalent to saying that firm 4 is 75.7 per cent efficient. As expected, the technical efficiency and the total factor productivity measurements are equal. To become fully efficient, under constant returns to scale, firm 4 must (radially) decrease input 1 from 133 units to $133 \times 0.757 = 100.7$ units, and must (radially) decrease input 2 from 189 units to $189 \times 0.757 = 143.0$ units.

Alternatively, the optimal quantities of inputs 1 and 2 may be found using the boundaries $100\lambda_1+90\lambda_2+150\lambda_3+133\lambda_4+152\lambda_5 = 100 \times 0.36 + 90 \times 0.718 = 100.7$ units for input 1 and $100\lambda_1+149\lambda_2+85.9\lambda_3+189\lambda_4+61\lambda_5 = 100 \times 0.36 + 149 \times 0.718 = 143.0$ units for input 2. Under constant returns to scale, the 115 units of output are technically attainable, using no less than 100.7 and 143.0 units of inputs 1 and 2, respectively.

*Slacks*

There are some cases where the radial reduction of inputs does not give the minimum technically feasible input vector for producing the specified output.

Figure 3.1 represents a piece-wise linear boundary of technology. Black dots represent the observed data of quantity of input 1 and 2 per unit of output for firms 1, 2, 3 and 4. The boundary of the technology is defined by the vertical line parallel to the $\frac{x_1}{y}$ axis, the line between firms 2 and 4, and the line parallel to the $\frac{x_2}{y}$ axis. Firms 1 and 3 are not efficient firms. To become technically efficient, firm 3 must reduce radially its input vector to point 3'. According to standard DEA, firm1 reduces radially its input vector to point 1'. But from Figure 3.1, firm 1 may perform an additional reduction of $\frac{x_1}{y}$ (with $\frac{x_2}{y}$ held constant) to point 2.

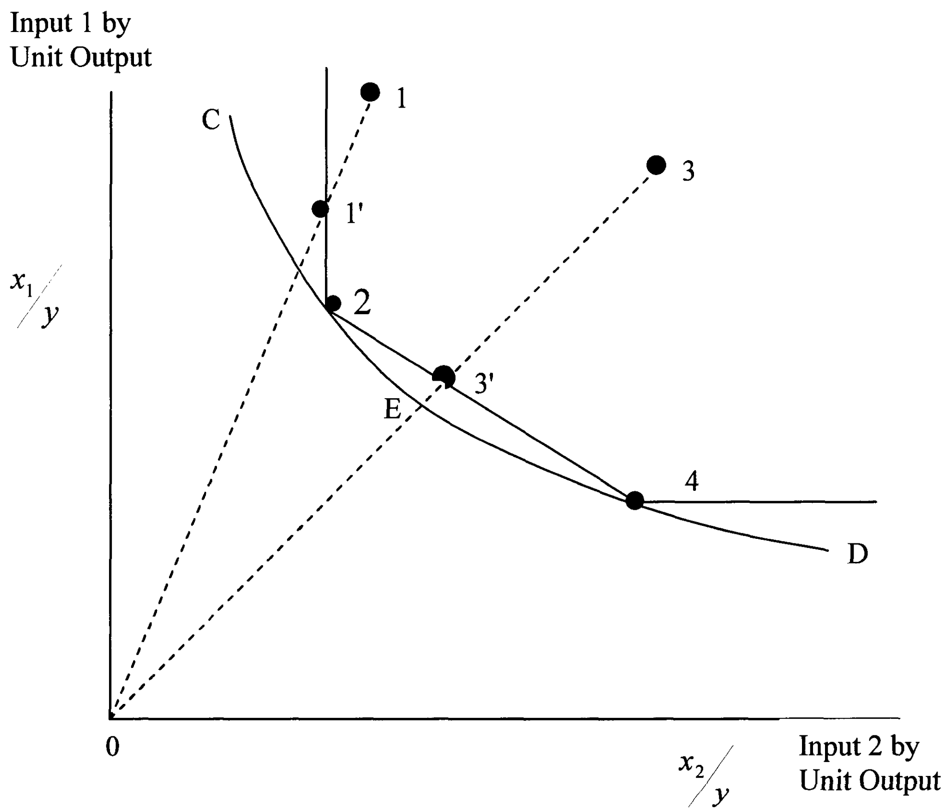


**Figure 3.1:** Input-Orientated Technical Efficiency Measurement
Constant Returns to Scale

The segment $\overline{1'2}$ is the additional reduction, and, in LP terminology, is referred to as slack. Coelli, Rao and Battese (1998, footnote p. 142) indicate that Koopmans (1951), (sic) "provides a more strict definition of technical efficiency which is equivalent to stating that a firm is only technically efficient if it operates on the frontier and furthermore that all associated slacks are zero".

In most cases, the technical efficiency given by DEA is larger than the technical efficiency given by a deterministic Cobb-Douglas production frontier. Consider curve CED of Figure 3.1, which is a deterministic Cobb-Douglas production frontier. Black dot points 2 and 4 are observed data and are DEA linear approximations to the production function. Firm 3 is radially projected onto point E over the Cobb-Douglas production frontier. The projection 03' is larger than 0E, the projection over the deterministic Cobb-Douglas production frontier. Then the technical efficiency given by DEA is larger than the technical efficiency given by a deterministic Cobb-Douglas production frontier.

*Output Orientated Technical Efficiency Measurement*

Consider now a set of K firms of an output-orientated system with constant returns-to-scale technology. Each firm produces J outputs, at quantities $y_k$, $k = 1, 2, \ldots, K$, using I inputs having quantities, $x_k$, $k = 1, 2, \ldots, K$,. For firm *e*, under output-orientation measurement of technical efficiency, equation (2.6.1) may be written using vector notation, as in equation (3.3.5), where the actual output quantity vector, $y_e$, equals the optimal output quantity vector, $y_e^*$, times the output technical efficiency measurement, $TE_{ie}$:

$$y_e = TE_{oe} \; y_e^* . \qquad (3.3.5)$$

With $e_e = (TE_{oe})^{-1}$, equation (3.3.5) may be written as in equation (3.3.6):

$$y_e^* = e_e \; y_e . \qquad (3.3.6)$$

The optimal quantities, $\boldsymbol{y}_e^*$, are unknown. DEA specifies $\boldsymbol{y}_e^*$ as a sum of weighted values of optimal quantities, $\boldsymbol{y}_k^*$, $k = 1, 2, \ldots,$ K. Under these conditions, with $\lambda_k$ being the weight of firm $k$, the right-hand side of equation (3.3.7) is the minimum feasible vector of input quantities, and the right-hand side of equation (3.3.8) is the maximum feasible vector of output quantities:

$$\boldsymbol{x}_e \geq \boldsymbol{X\lambda} \tag{3.3.7}$$

$$e_e\, \boldsymbol{y}_e \leq \boldsymbol{Y\lambda} \tag{3.3.8}$$

The vector $\boldsymbol{y}_e$ is fixed and known. To maximise $\boldsymbol{y}_e^*$, the value of $e_e$ has to be a maximum. The LP to maximise $e_e$ may be written as:

$$\text{maximise } e_e \tag{3.3.9}$$

subject to:

$$\boldsymbol{x}_e \geq \boldsymbol{X\lambda}$$

$$e_e\, \boldsymbol{y}_e \leq \boldsymbol{Y\lambda}$$

$$TE_{oe} = 1/e_e$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0}.$$

The main interest is to determine the values of the weights $\boldsymbol{\lambda}$ that, maximising $e_e$, minimise $TE_{oe}$. Equations (3.3.7) and (3.3.8) set constraints for $\boldsymbol{\lambda}$, because the components of vectors, $\boldsymbol{x}_e$, $TE_{oe}$ and of matrix $\boldsymbol{X}$ contain observed data.

In problem (3.3.9), the subscript, $e$, stands for the firm under evaluation, $e = 1, 2, \ldots,$ K. For this reason, to evaluate the technical efficiency of the K firms requires the solution of problem (3.3.9) for each of the K firms.

To illustrate these concepts, consider the data of Table 3.1 and for each firm determine the technical efficiency, $TE_{oe}$, as defined with problem (3.3.9) and with $e_0 = (TE_{oe})^{-1}$.

The LP problems for firms 1 and 4, for the output-orientated system and constant returns-to-scale technology are defined by:

| Firm 1: | Firm 4: |
|---|---|
| maximise $e_1$ | maximise $e_4$ |
| subject to: | subject to: |
| $100\lambda_1+110\lambda_2+120\lambda_3+115\lambda_4+103\lambda_5 \geq 100\,e_1$ | $100\lambda_1+110\lambda_2+120\lambda_3+115\lambda_4+103\lambda_5 \geq 115\,e_1$ |
| $100\lambda_1 +90\lambda_2 +150\lambda_3+133\lambda_4 +152\lambda_5 \leq 100$ | $100\lambda_1 +90\lambda_2 +150\lambda_3+133\lambda_4 +152\lambda_5 \leq 133$ |
| $100\lambda_1 +149\lambda_2 +85.9\lambda_3+189\lambda_4 +61\lambda_5 \leq 100$ | $100\lambda_1 +149\lambda_2 +85.9\lambda_3 +189\lambda_4 +61\lambda_5 \leq 189$ |
| $\lambda_1, \lambda_2, \lambda_3,, \lambda_4, \lambda_5 \geq 0$ | $\lambda_1, \lambda_2, \lambda_3,, \lambda_4, \lambda_5 \geq 0$ |

Similar LP problems are required for firms 2, 3 and 5. Table 3.4 shows the solution for the five firms.

For constant returns-to-scale technology, the output-orientated technical efficiency measurement is the same as that determined under input orientation. Nonetheless, as in the case of input orientation, firms 1 and 2 are the technically efficient firms, whose weighted output quantities determine the optimal output quantity of firm 4. For firm 4, the optimal output is $100\lambda_1+110\lambda_2+120\lambda_3 +115\lambda_4 +103\lambda_5$ = optimal output, $y_0^*$. Replacing $\lambda_1 = 0.4758$ and $\lambda_2 = 0.9492$, $y_e^* = 152$ units. Alternatively, this optimal quantity may be found as $y_e^* = e_e\, y_k$; replacing $y_e^* = 1.322 \times 115 = 152$ units.

**Table 3.4:** Solution to Problem (3.3.9) for the Five Firms: Output-Orientated Technical Efficiency Measurement and Constant Returns-to-Scale Technology

| Firm | $e_k$ | Technical Efficiency | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|---|
| 1 | 1.0 | 1.0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 1.0 | 1.0 | 0 | 1 | 0 | 0 | 0 |
| 3 | 1.0 | 1.0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 1.322 | 0.757 | 0.4758 | 0.9492 | 0 | 0 | 0 |
| 5 | 1.0 | 1.0 | 0 | 0 | 0 | 0 | 1 |

It may be observed that the values of $\lambda_k$ differ from input to output orientation. To show that lambdas for input- and output-orientation measurement of technical efficiency are different, consider Figure 3.2, which presents a system of one input and one output. The line 0CA represents the boundary of technology with constant returns to scale.

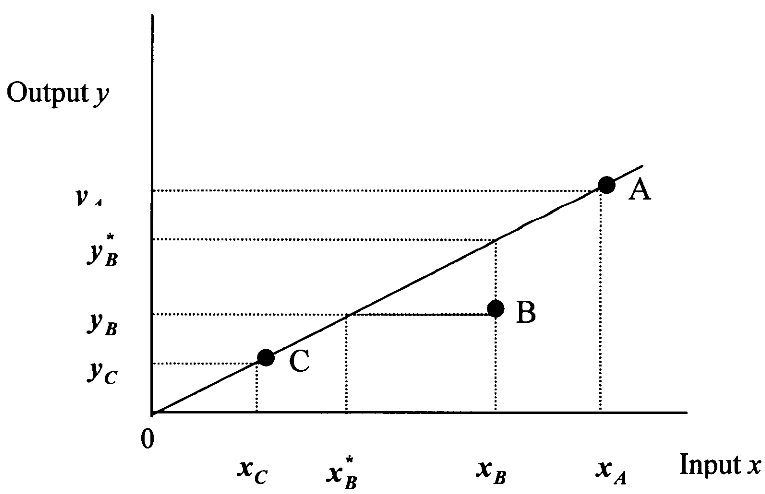


**Figure 3.2:** $\lambda_k$ Depends on the System's Orientation

Firms A and C are technically efficient. Firm B is not efficient and produces $y_B$ units of output using $x_B$ units of input.

For the input-orientated system, to be technically efficient, firm B must reduce input quantity $x_B$ to the minimum technically feasible $x_B^*$ for the same output, $y_B$. At this optimal input-output condition, the input, $x_B^*$, is a weighted combination of optimal inputs of firms A and C. If $\lambda_{iA}$ and $\lambda_{iC}$ are the respective weights, then $x_B^* = x_A\lambda_{iA} + x_C\lambda_{iC}$. Similarly, the output, $y_B$, is a weighted combination of optimal outputs of firms A and C, then $y_B = y_A\lambda_{iA} + y_C\lambda_{iC}$. The values of $\lambda_{iA}$ and $\lambda_{iC}$ define the solution of these two equations.

$$\begin{bmatrix} x_B^* \\ y_B \end{bmatrix} = \begin{bmatrix} x_A & x_C \\ y_A & y_C \end{bmatrix} \begin{bmatrix} \lambda_{iA} \\ \lambda_{iC} \end{bmatrix} \qquad (3.3.10)$$

For the output-orientated system, to be technically efficient, firm B must expand output quantity, $y_B$, to the maximum technically feasible, $y_B^*$, for the same input, $x_B$. At this optimal input-output condition, the output $y_B^*$ is a weighted combination of optimal inputs of firms A and C. If $\lambda_{oA}$ and $\lambda_{oC}$ are the respective weights, then $y_B^* = y_A\lambda_{oA} + y_C\lambda_{oC}$. Similarly, the input, $x_B$, is a weighted combination of optimal outputs of firms A and C, then $x_B = x_A\lambda_{oA} + x_C\lambda_{oC}$. The values of $\lambda_{oA}$ and $\lambda_{oC}$ define the solution of these two equations.

$$\begin{bmatrix} x_B \\ y_B^* \end{bmatrix} = \begin{bmatrix} x_A & x_C \\ y_A & y_C \end{bmatrix} \begin{bmatrix} \lambda_{oA} \\ \lambda_{oC} \end{bmatrix} \qquad (3.3.11)$$

Because equations (3.3.10) and (3.3.11) have the same square matrix and different left-hand side vector, the values of $\lambda_{iA}$ and $\lambda_{oA}$ are different; the same for the value of $\lambda_{iC}$ and $\lambda_{oC}$, as previously stated.

Another important issue is that, under constant returns-to-scale technology, the input- and output-orientated efficiency measurements are equal (Cooper, Seiford and Tone, 2000, Section 3.8). Consider Figure 3.3, where line 0Q represents the constant returns-to-scale technology for a system of one output $y$ and one input $x$. Point P represents a non-efficient firm that produces $y_1$ units of product, using $x_1$ units of input.

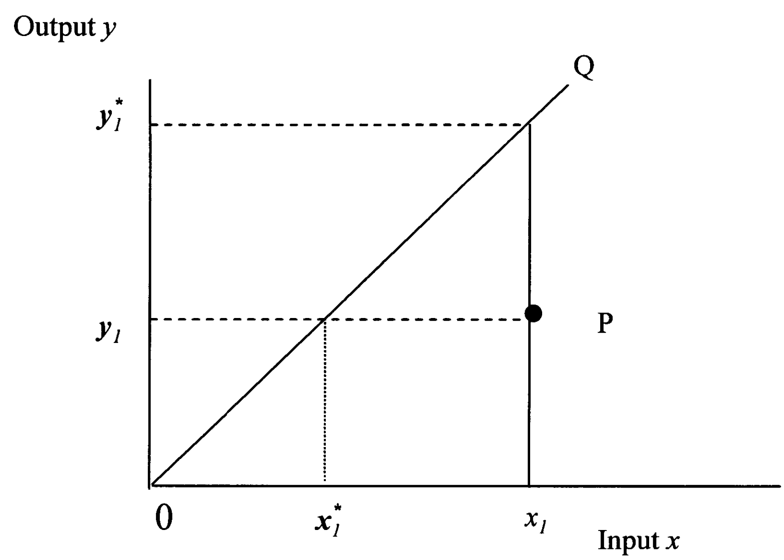


**Figure 3.3:** Input- and Output-Orientated Technical Efficiency Measurement

To be technically efficient, firm P must reduce input $x_1$ to the minimum technically feasible, $x_1^*$, for the same output, $y_1$. This is the input-orientated case, and the input-orientated technical efficiency measurement is $TE_i = \dfrac{0x_1^*}{0x_1}$.

Instead of reducing input, firm P may expand output $y_1$ to the maximum technically feasible, $y_1^*$, for the same input, $x_I$. This is the output-orientated case, and the output-orientated technical efficiency measurement is $TE_0 = \frac{0y_I}{0y_1^*}$. But $\frac{0y_1}{0x_I^*} = \frac{0y_1^*}{0x_1}$ so that, by re-arranging the terms, we obtain $\frac{0y_1}{0y_1^*} = \frac{0x_1^*}{0x_1}$ and so $TE_0 = TE_i$.

Although the primal and the dual models give the same information, the dual optimisation problem involves fewer constraints than the primal and, hence, it is computationally more efficient. Additionally, for the purpose of this thesis, the dual model is better suited than the primal, because the former has the boundary of the technology in explicit form. The power of this model is that the boundary of technology may be used with any objective function of interest. Another important issue of the dual model is that the weight of firms (the lambdas) is dimensionless. This means that input and output quantities may be expressed in the most convenient set of units, i.e., AU$ or US$.

So far, we have only dealt with constant returns to scale. The following section analyses variable returns to scale.

*Target Assignment and Peers Recognition*

In this chapter, specific DEA formulations are derived for determining the optimal quantity of inputs (or outputs) that optimise the technical efficiency measurement. In the following chapters, two terms are referred frequently: *targets* and *peers*.

A *target* is the optimal quantity of inputs or outputs that a firm has to achieve to optimise a function of interest. In previous sections, the objective functions have a common set of constraints in inputs and outputs. These constraints define the boundaries of the technology. The transformation function or the boundary of the technology is implicit in these constraints. The common constraints to the different optimisation problems are:

$$\mathbf{x}_e^* \geq \mathbf{X}\boldsymbol{\lambda} \qquad (3.3.12)$$

$$\mathbf{y}_e^* \leq \mathbf{Y}\boldsymbol{\lambda}$$

$$\boldsymbol{\lambda} \geq \mathbf{0},$$

where $\mathbf{x}_e^*$ and $\mathbf{y}_e^*$ are the optimal input and output quantity vectors, respectively.

Inspecting the solution to numerical problems presented in previous sections, it is apparent that for each firm under evaluation, there is a set of $\lambda s$ equal to zero, and other set of $\lambda s$ different from zero. Once one has solved the technical efficiency measurement problem for each of the K firms, those firms that have their $\lambda$ s different from zero, at least once, are 100 per cent technologically efficient; they are the best-performing firms, and may be considered as potential role models to non-efficient firms.

*Peers* of firm *e* are those technically efficient firms that exhibit $\lambda$s larger than zero, when solving the optimisation problem for firm *e*. A best-performing firm, *e*, has itself as the unique peer, and $\lambda_e = 1$.

The firms that have their $\lambda$ equal to zero, for the K times that the optimisation problem is solved, are non-efficient firms.

For input-orientated systems, while keeping constant the output, $\boldsymbol{y}_e$, the optimal set of inputs is:

$$\boldsymbol{x}_e^* = \boldsymbol{X}\boldsymbol{\lambda} \tag{3.3.13}$$

This means that $\boldsymbol{x}_e^*$ is a linear combination of its peers. The weight of each peer is its respective $\lambda$.

For output-orientated systems, while keeping constant the input, $\boldsymbol{x}_e$, the optimal set of outputs is:

$$\boldsymbol{y}_e^* = \boldsymbol{Y}\boldsymbol{\lambda} \tag{3.3.14}$$

This means that $\boldsymbol{y}_e^*$ is a linear combination of its peers. The weight of each peer is its respective $\lambda$.

For the purpose of this thesis, the selection of DEA is based on its advantages over parametric methods for describing the transformation function or the boundary of the technology. Nonetheless, DEA has some limitations that must be considered in empirical applications. Section 3.6 presents some limitations of DEA.

### 3.4 Returns to Scale

This section analyses returns to scale and presents the concept of variable returns to scale, complementing the previous explicit assumption that the boundary of the technology exhibits constant returns to scale. This section includes how to identify the kind of returns to scale of a boundary of technology, and includes the addendum to the models presented so far to assess performance under variable returns to scale. As is mentioned in Chapter 2, the returns to scale is a property of the boundary of technology, not of the firm (Thanassoulis, 2001, p. 124).

Consider a one-input, one-output system. Firm $e$ of this system performs at 100 per cent technical efficiency, using $x_e^*$ units of input to produce $y_e^*$ units of output. The boundary of technology of this system behaves with constant returns to scale if another firm on the boundary uses $\alpha x_e^*$ units of input to produce $\alpha y_e^*$ units of output, $\alpha > 0$.

As an example, consider the college system of a university. This system considers the number of student members of the college as the sole output and the operating expenditure as the sole input.

Assume that the boundary of technology of this system behaves with constant returns to scale. Assume that College E.P. and College S.T. are members of this system and that they perform 100 per cent technically efficiently. Then, if College S.T. has half the members that College E.P. has, it should expend half the operating expenditure of College E.P. Nonetheless, there are two other cases:

Increasing the number of students, within certain limits, may make it possible to control the increase of operating expenditure at a lower rate than increases the number of students. This is the increasing returns-to-scale case.

But increasing the number of students, beyond the limits mentioned above, may require one to increase the operating expenditure at a higher rate than increases the number of students. This is the decreasing returns-to-scale case.

In general terms, a technology exhibits increasing returns to scale if a proportional increase of input quantities gives a more than proportional increase of output quantities. A technology exhibits constant returns to scale if a proportional increase of input quantities gives the same proportional increase of output quantities. A technology exhibits decreasing returns to scale if a proportional increase of input quantities gives a less than proportional increase of output quantities.

A technology may exhibit the three types of returns to scale on different parts of the frontier surface. For this reason, scale efficiency (SE) is measured to provide an insight in the firm's performance with respect to the optimal input-output vectors. *Scale efficiency* is defined as the ratio of the technical efficiency measurement that the firm should have if the system operates under constant returns-to-scale technology, over the technical efficiency measurement of the firm under the real returns to scale of the system (in the area of the frontier where the firm is operating).

To illustrate these concepts, consider Figure 3.4, which presents a system of one input and one output. The black dot represents firm *e* that produces $y_0$ units of output using $x_0$ units of input. Straight line 0F represents a technology that exhibits constant returns to scale.

The variable returns-to-scale technology, represented by lines ABCDE, exhibits the three analysed types of returns to scale. Line AB represents that part of the technology that exhibits increasing returns to scale, line BC represents that part of the technology that exhibits constant return to scale, and line CDE represents that part of the technology that exhibits decreasing returns to scale.

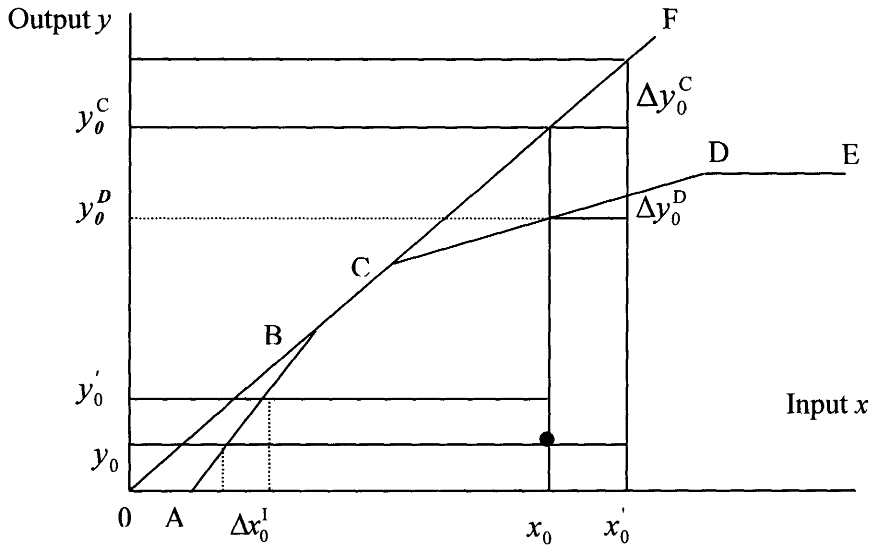


**Figure 3.4:** Boundaries of Technology Exhibits Constant, Increasing and Decreasing Returns to Scale.

For the output-orientated case, to be technically efficient, firm $e$ should expand output to the maximum feasible. From Figure 3.4, the maximum feasible expansion is limited to $y_0^D$ by the boundary of decreasing returns to scale. The output-orientated technical efficiency of Firm $e$ is $\frac{y_0}{y_0^D}$, as previously stated. Under constant returns to scale, firm $e$ should expand output to $y_0^C$ with technical efficiency $\frac{y_0}{y_0^C}$. The output-orientated scale efficiency of firm $e$ is the ratio $\frac{y_0^{\mathrm{D}}}{y_0^{\mathrm{C}}}$. Correspondingly, there is an input-orientated scale efficiency measurement.

Consider now that firm $e$ increases its input from the actual value $x_0$ to $x_0'$. From Figure 3.4, on the boundary of the decreasing returns-to-scale technology, the optimal value of output increases by $\Delta y_0^{\mathrm{D}}$. If the slope of the segment CD decreased, then the increment of output, $\Delta y_0^{\mathrm{D}}$, would decrease. From Figure 3.4, on the boundary of the constant return-to-scale technology, the optimal value of output increases by $\Delta y_0^{\mathrm{C}}$. This output increment is larger than the output increment under decreasing returns to scale.

Consider now that firm $e$ increases its output from its actual value $y_0$ to $y_0'$. From Figure 3.4, on the boundary of the increasing returns to scale part of the technology, the optimal value of input increases by $\Delta x_0^{\mathrm{I}}$. If the slope of the segment AB increased, then the increment of input, $\Delta x_0^{\mathrm{I}}$, would decrease. This input increment is smaller than the input increment under constant returns to scale.

For multiple-input multiple-output systems, the definition of increasing, constant and decreasing returns to scale may be generalised as follows (Thanassoulis, 2001, p. 125):

Let firm $e$ be 100 per cent technically efficient. This firm produces $\boldsymbol{y}_0^*$ units of output, using $\boldsymbol{x}_0^*$ units of input. Let us assume that increasing inputs to $\alpha \boldsymbol{x}_0^*$, the outputs increase to $\beta \boldsymbol{y}_0^*$, and that the firm remains 100 per cent technically efficient. The net increase of outputs is $\beta \boldsymbol{y}_0^* - \boldsymbol{y}_0^* = (\beta-1)\boldsymbol{y}_0^*$ and the net increase of inputs is $(\alpha-1)\boldsymbol{x}_0^*$.

Defining $\rho^* = \boldsymbol{y}_0^*/\boldsymbol{x}_0^*$ and $\rho = (\beta-1)\boldsymbol{y}_0^*/(\alpha-1)\boldsymbol{x}_0^*$, then $\frac{\rho}{\rho^*} = \frac{(\beta-1)}{(\alpha-1)}$. As $\alpha \to 1$, the ratio $\rho/\rho^*$ measures the local radial rate of change of outputs to local radial change of inputs; it measures the local returns of scale. As stated above, if $\lim_{\alpha\to 1}\frac{(\beta-1)}{(\alpha-1)} > 1$ then the technology exhibits local increasing returns to scale at $(\boldsymbol{x}_0^*, \boldsymbol{y}_0^*)$; if $\lim_{\alpha\to 1}\frac{(\beta-1)}{(\alpha-1)} = 1$ then the technology exhibits local constant returns to scale at $(\boldsymbol{x}_0^*, \boldsymbol{y}_0^*)$; if $\lim_{\alpha\to 1}\frac{(\beta-1)}{(\alpha-1)} < 1$ then the technology exhibits local decreasing returns to scale at $(\boldsymbol{x}_0^*, \boldsymbol{y}_0^*)$.

To formalise the constraint that defines variable returns to scale, consider problem (3.3.4) for efficient firm 3:

Minimise $TE_{i3}$,

subject to:

$100\lambda_1 + 110\lambda_2 + 120\lambda_3 + 115\lambda_4 + 103\lambda_5 \geq 120$

$100\lambda_1 + 90\lambda_2 + 150\lambda_3 + 133\lambda_4 + 152\lambda_5 \leq 150\ TE_{i3}$

$100\lambda_1 + 149\lambda_2 + 85.9\lambda_3 + 189\lambda_4 + 61\lambda_5 \leq 85.9\ TE_{i3}$

$\lambda_1, \lambda_2, \lambda_3,, \lambda_4, \lambda_5 \geq 0.$

For this particular problem, let us specify $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = sl_3 > 0$, then the inputs and outputs of the three constraints that define the boundary of technology may be scaled by division of respective quantities by $sl_3$. Doing this, we have

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1.$$

Under constant returns to scale, optimal sets of inputs and outputs remain optimal when scaled. For this reason, $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1$ is not included as an extra constraint to the optimisation problems. Conversely, under variable returns to scale, the extra constraint, $\mathbf{1}'\lambda=1$, must be included, because under variable returns to scale, optimal sets of inputs and outputs do not remain optimal when scaled.

Nevertheless, and as stated previously, when measuring the technical efficiency for an efficient firm, the value of $\lambda$ for this firm is one and zero for the rest of the firms. In this particular case, and under constant returns to scale, the summation of lambdas also equals one.

In terms of benchmarking, the sum of lambdas has the following meanings:

$\mathbf{1}'\lambda \leq 1$, non-increasing returns to scale; the firm under study is not benchmarked against firms substantially larger than it, but may be compared with firms smaller than it.

$\mathbf{1}'\lambda=1$, variable returns to scale; the firm under study is benchmarked against firms of similar size. The firm under study is optimised as a convex combination of comparable observed firms.

$\mathbf{1}'\lambda$ not constrained, the firm under study is benchmarked against any other. If the firms are smaller than it, the sum will be larger than 1.0; if firms are larger than it, the sum will be smaller than 1.0; and if there is a mix of larger and smaller firms, the sum may be any number.

Following Banker, Charnes and Cooper (1984), under variable returns-to-scale technology, problem (3.3.4) becomes

$$\text{Minimise } TE_{ie} \qquad (3.4.1)$$

subject to:

$$\boldsymbol{X\lambda} \leq TE_{ie}\ \boldsymbol{x}_e$$

$$\boldsymbol{Y\lambda} \geq \boldsymbol{y}_e$$

$$\mathbf{1}'\boldsymbol{\lambda}=1$$

$$\boldsymbol{\lambda} \geq \mathbf{0}.$$

With the constraint, $\mathbf{1}'\boldsymbol{\lambda}=1$, the hyperplanes that define the boundary of technical efficiency deflect, allowing interpolation of smaller hyperplanes. These smaller hyperplanes are closer to non-efficient firms than the hyperplanes are under constant returns to scale. For this reason, the technical efficiency of a non-efficient firm, measured under variable returns to scale, is higher than measured under constant returns to scale.

As an example, consider data of Table 3.1 and determine the technical efficiency for the five firms, as defined with problem (3.4.1) for variable returns-to-scale technology. The LP problems for firms 1 and 4, for the input-orientated system and variable returns-to-scale technology, are defined in the box below. Similar LP problems are required for firms 2, 3 and 5. Table 3.5 shows the solution for the five firms.

| Firm 1: | Firm 4: |
|---|---|
| minimise $TE_{i1}$ | minimise $TE_{i4}$ |
| subject to: | subject to: |
| $100\lambda_1 + 90\lambda_2 + 150\lambda_3 + 133\lambda_4 + 152\lambda_5 \leq 100\ E_1$ | $100\lambda_1 + 90\lambda_2 + 150\lambda_3 + 133\lambda_4 + 152\lambda_5 \leq 133\ TE_4$ |
| $100\lambda_1 + 149\lambda_2 + 85.9\lambda_3 + 189\lambda_4 + 61\lambda_5 \leq 100\ E_1$ | $100\lambda_1 + 149\lambda_2 + 85.9\lambda_3 + 189\lambda_4 + 61\lambda_5 \leq 189\ TE_4$ |
| $100\lambda_1 + 110\lambda_2 + 120\lambda_3 + 115\lambda_4 + 103\lambda_5 \geq 100$ | $100\lambda_1 + 110\lambda_2 + 120\lambda_3 + 115\lambda_4 + 103\lambda_5 \geq 115$ |
| $\lambda_1 + \lambda_2 + \lambda_3, + \lambda_4 + \lambda_5 = 1$ | $1\lambda_1 + \lambda_2 + \lambda_3, + \lambda_4 + \lambda_5 = 1$ |
| $\lambda_1, \lambda_2, \lambda_3,, \lambda_4, \lambda_5 \geq 0$ | $\lambda_1, \lambda_2, \lambda_3,, \lambda_4, \lambda_5 \geq 0$ |

**Table 3.5:** Solution to Problem (3.4.1) for the Five Firms: Input-Orientated Technical Efficiency and Variable Returns-to-Scale Technology

| Firm | Technical Efficiency | $\lambda_1$ | $\lambda_2$ | $\lambda_3,$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|---|---|
| 1 | 1.0 | 1 | 0 | 0 | 0 | 0 |
| 2 | 1.0 | 0 | 1 | 0 | 0 | 0 |
| 3 | 1.0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0.902 | 0 | 0.50 | 0.50 | 0 | 0 |
| 5 | 1.0 | 0 | 0 | 0 | 0 | 1 |

Again, following Thanassoulis (2001, p. 93), the information obtained from this DEA model is:

Firms 1, 2, 3 and 5 are fully efficient, having the maximum attainable efficiency.

Firm 4 is not efficient, having a relative efficiency of 0.902. This is equivalent to saying that firm 4 is 90.2 per cent efficient. As expected under variable returns to scale, firm 4 is more efficient than under constant returns to scale. Under constant returns to scale technology, firm 4 is 75.7 per cent efficient. The scale efficiency of firm 4 is 75.7/90.22 = 0.839

To be fully efficient, under variable returns to scale, firm 4 must decrease input 1 from 133 units to $100\lambda_1+90\lambda_2+150\lambda_3+133\lambda_4+152\lambda_5$ = 120 units and must decrease input 2 from 189 units to $100\lambda_1+149\lambda_2+85.9\lambda_3+189\lambda_4+61\lambda_5$ = 117.45 units. Operating at these optimal conditions, if firm 4 increases each input by 3.0 per cent ($\alpha = 1.03$), to remain technically efficient it must increase output by only 0.52 per cent ($\beta$=1.0052). Because $\rho/\rho^* = 0.173$, the technology exhibits decreasing returns to scale in the neighbourhood of $x_{1k}^*$=120 units and $x_{2k}^*$=117.45 units.

### 3.5 Optimising Behaviour and the DEA Dual Formulation

In previous sections, the focus is on the dual DEA formulation of relative productivity and technical efficiency measurements. In this section, the fundamental structure of dual DEA formulations, presented in Section 3.3, is applied to formalise the optimising behaviour of three possible firms. By fundamental structure we mean the expression for the boundary of technology that constrains the upper limits of the output quantities and the lower limits of the inputs quantities to be used for producing those outputs.

The three possible optimising behaviours, profit maximisation, cost minimisation and revenue maximisation, yield the same solutions, as presented in Section 2.4.

In general terms, under certainty, firms seek to maximise their profits (Chambers, 1988, p. 120).[7] For the purposes of this thesis, we assume the accounting sense of the term, *net profit,* as the residual after deduction of all money costs, i.e., sales revenue minus wages, salaries, rents, raw materials, etc. (Bannock, Baxter and Davis, 1998, p. 335). Given that the production technology is described by a constant returns-to-scale DEA frontier, we can restate the profit maximisation objective, defined in problem (2.4.1), as:

$$\pi_e(\boldsymbol{p},\boldsymbol{w}) = \max_{\boldsymbol{x},\boldsymbol{y}} (\boldsymbol{yp}-\boldsymbol{xw}) \qquad (3.5.1)$$

subject to:

$$\boldsymbol{x}_e^* \geq \boldsymbol{X\lambda}$$

$$\boldsymbol{y}_e^* \leq \boldsymbol{Y\lambda}$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0},$$

where $\boldsymbol{p}$ is the $(1\times J)$ vector of prices for J outputs; $\boldsymbol{y}$ is the $(J\times 1)$ output quantity vector; $\boldsymbol{w}$ is the $(1\times I)$ vector of input prices; $\boldsymbol{x}$ is the $(I\times 1)$ input quantity vector; UI is the number of inputs; $\boldsymbol{\lambda}$ is the $(K\times 1)$ vector of weights for K firms; $\boldsymbol{X}$ is the $(I\times K)$ matrix of input quantity vectors; $\boldsymbol{Y}$ is the $(J\times K)$ matrix of output quantity vectors; and $\boldsymbol{x}_e^*$ and $\boldsymbol{y}_e^*$ are the optimal input and output vectors, respectively, for firm *e*.

In problem (3.5.1), the constraints, $\boldsymbol{x}_e^* \geq \boldsymbol{X\lambda}$, $\boldsymbol{y}_e^* \leq \boldsymbol{Y\lambda}$, and $\boldsymbol{\lambda} \geq \boldsymbol{0}$, define the technical boundary for this system.

[7] For the purpose of this thesis, we do not consider utility maximisation.

As mentioned in Section 2.4, Färe, Grosskopf and Lovell, (1994, p. 213) point out that there may be no finite solution to problem (3.5.1). In fact, under constant returns to scale, if no additional constraints are imposed to problem (3.5.1), there is no bounded solution. The reason for this is that as long as ***py-wx*** is positive, $y$ increases without bound to maximise the profit, $\pi_e(\boldsymbol{p},\boldsymbol{w})$. Unbounded $y$ drives unbounded $x$. Nonetheless, under decreasing return-to-scale technology, the output is limited to some value, as is illustrated by boundary CDE of Figure 3.4. With bounded output, profit maximisation is feasible. A decreasing returns-to-scale technology constraint must be added to problem (3.5.1) for it to have a solution.

From this point on, we refer to problem (3.5.1) as the *fundamental dual DEA problem* or, simply, the fundamental problem, because from this problem we can derive the DEA problems that optimise profit, revenue and cost. These three optimisation problems are of interest for this thesis.

*Profit Maximisation; Variable Returns-to-scale Technology*

Consider the data of Table 3.6 and determine the optimal output vector, $\boldsymbol{y}_e^*$ and optimal input vector, $\boldsymbol{x}_e^*$, that maximise profit, $\pi_e$, under variable returns-to-scale technology. Then, maximum $\pi_e(\boldsymbol{p},\boldsymbol{w})$ = maximum $(\boldsymbol{yp}-\boldsymbol{xw})$ and the problem is similar to problem (3.5.1), with the additional constraint that defines variable-returns-to-scale technology:

$$\mathbf{1}'\lambda=1 \qquad (3.5.2)$$

The variable returns-to-scale constraint, $\mathbf{1}'\lambda=1$, constrains the value of the weights such that the output quantity vector has an upper limit. This upper limit for outputs limits the increase of the input quantity vector, and the profit maximisation problem has a bounded solution.

However, under a constant returns-to-scale technology, we can avoid the unbounded expansion of outputs and inputs, by considering the following special cases (Chambers, 1988, p. 121).

The first case is a short-run optimisation that involves maximising profit for a given fixed output vector, $y$. With this constraint, profit maximisation corresponds to a cost minimisation problem.

The second case is a long-run optimisation that involves maximising profit for a given fixed input vector, $x$. With this constraint, profit maximisation corresponds to a revenue maximisation problem.

*Cost of Inputs Minimisation; Constant Returns-to-scale Technology*

For fixed output vector, $y$, the main purpose is to determine the optimal input quantities, $x^*$, that maximises profit, $\pi_e$ (i.e., minimising cost of inputs). The cost-minimizing problem is (Thanassoulis, 2001, p. 81):

minimise $(\boldsymbol{xw})$ (3.5.3)

subject to:

$\boldsymbol{x}_e^* \geq \boldsymbol{X}\lambda$

$\boldsymbol{y}_e \leq \boldsymbol{Y}\lambda$

$\boldsymbol{\lambda} \geq \mathbf{0}$ ,

where $\boldsymbol{x}_e^*$ is the vector of optimal inputs. In problem (3.5.3), the constraints, $\boldsymbol{x}_e^* \geq \boldsymbol{X}\boldsymbol{\lambda}$ , $\boldsymbol{y}_e \leq \boldsymbol{Y\lambda}$, and $\boldsymbol{\lambda} \geq \mathbf{0}$, define the boundary of technical efficiency for this system. The optimal solution vector is denoted by $\boldsymbol{x}_e^*$ to indicate that it is on the boundary of technical efficiency. If prices are different from firm to firm, the cost efficient peer may be different from firm to firm.

Consider Figure 3.5 that shows the effect of the input price in the cost efficient peer determination. The observed data define the isoquant ABCD, where A, B, C, and D are technically efficient firms. Black dot firm E is a price inefficient firm. Assume that input prices may be different for each firm. Let be $w_1$ and $w_2$ be the prices of input 1 and input 2 for firm E. The isocost line II' represents the total cost of inputs for firm E, TC = $w_1 x_1 + w_2 x_2$ .

By parallel translation of that isocost line to the frontier of the technology, the minimum total cost of inputs is at the point where the isocost is tangent to the boundary of technology. In Figure 3.5, the isocost line T1 is tangent at the boundary of the technology at the input quantities per unit output of firm B. At that point, the slope of the isocost line T1 (and II', because the lines are parallel) is (- $w_2 / w_1$ ).

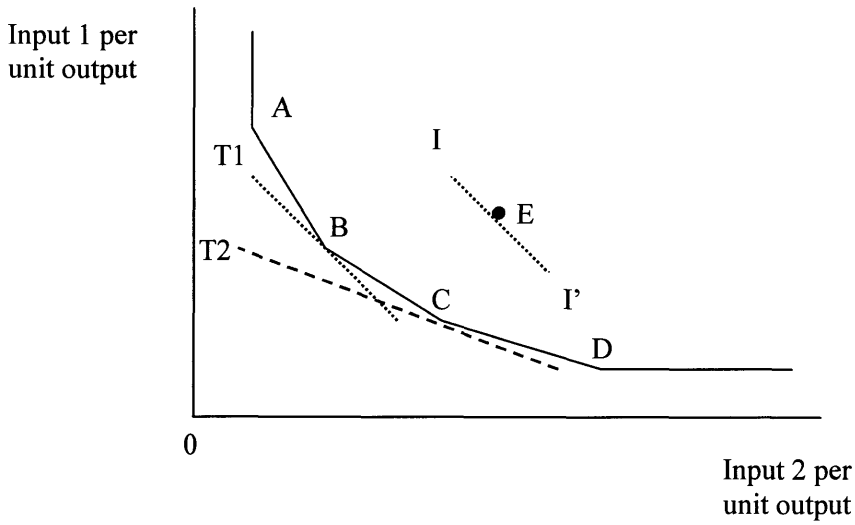


**Figure 3.5:** Effect of Input Prices in Input-Price Efficient Firm

If input prices for firm E change to $w_1^{'}$ and $w_2^{'}$ then the minimum total cost of input may be at firm C. The isocost line tangent at C, T2, has slope ($-w_2^{'}/w_1^{'}$). If the input prices of firm E are such that the slope of the isocost line is the same as the slope of boundary of technology that is defined by firms B and C, then there are infinite solutions along the line BC. In that case, managers may choose the optimal input quantities considering additional factors, i.e., lower capital investment, employment policy, environmental impact of inputs mixture, supply reliability, etc.

*Cost of Inputs Minimisation; Variable Returns-to-scale Technology*

For fixed output vector, $\boldsymbol{y}$, the main purpose is to determine the optimal input quantities, $\boldsymbol{x}^*$, that maximise profit, $\pi_e$ (i.e., minimising cost of inputs). The problem is similar to (3.5.1), with the additional constraint, $\mathbf{1}'\boldsymbol{\lambda}=1$, that defines variable-returns-to-scale technology:

$$\mathbf{1}'\boldsymbol{\lambda}=1. \tag{3.5.4}$$

The optimal solution vector is denoted by $\boldsymbol{x}_e^*$ to indicate that it is on the isoquant, on the boundary of technical efficiency.

*Revenue Maximisation; Constant Returns-to-scale Technology*

Consider now a fixed input vector $\boldsymbol{x}$, and that the main purpose is to determine the quantities of outputs, $\boldsymbol{y}_e^*$, that maximises the profit $\pi_e$, (i.e., maximising revenue). Then, maximum $\pi_e\,(\boldsymbol{p},\boldsymbol{w})$ = maximum $(\boldsymbol{yp})$ and the problem is (Thanassoulis, 2001, p. 82):

$$\text{maximise } (\boldsymbol{yp}) \qquad (3.5.5)$$

subject to:

$$\boldsymbol{x}_e \geq \boldsymbol{X}\boldsymbol{\lambda}$$

$$\boldsymbol{y}_e^* \leq \boldsymbol{Y}\boldsymbol{\lambda}$$

$$\boldsymbol{\lambda} \geq \boldsymbol{0}.$$

In problem (3.5.5), the constraints, $\boldsymbol{x}_e \geq \boldsymbol{X}\boldsymbol{\lambda}$, $\boldsymbol{y}_e^* \leq \boldsymbol{Y}\boldsymbol{\lambda}$, and $\boldsymbol{\lambda} \geq \boldsymbol{0}$, define the boundary of technical efficiency for this system. The optimal solution vector is denoted by $\boldsymbol{y}_e^*$ to indicate that it is on the isoquant, on the boundary of technical efficiency

*Revenue Maximisation; Variable Returns-to-scale Technology*

As before, determine the optimal output vector, $\boldsymbol{y}_e^*$, that maximises the profit, $\pi_e$, for a variable returns-to-scale technology. The problem is similar to (3.5.5), with the additional constraint, $\boldsymbol{1}'\boldsymbol{\lambda}=1$, that defines variable-returns-to-scale technology:

$$\boldsymbol{1}'\boldsymbol{\lambda}=1 \qquad (3.5.6)$$

In this section, we have established the foundations of the DEA models, which we build upon to provide a new model that considers optimal paths of adjustment. The basic intertemporal model is developed in Chapter 4. The basic model includes, in implicit form, the boundary of the technology, such that the objective function is the present value of cash flows over a number of time periods.

### 3.6 Limitations of DEA

In the Section 3.1, we noted that the use of a parametric method for data analysis assumes that an econometric model defines the data generating process involved. However, the selected functional form may be misspecified. DEA, as a non-parametric method, overcomes this limitation.

Nonetheless, DEA has some limitations. Coelli, Rao and Battese (1998, pp. 180-181) present a list of "limitations and possible problems that one may encounter in conducting a DEA". Of that list, the following five points are pertinent for this thesis.

Measurement errors may influence the shape and position of the frontier of the technology.

Outliers are an example of possible measurement errors.

The exclusion of relevant inputs or outputs may determine useless *frontiers* of the technology.

Not including environmental variables may give misleading indications of relative managerial competence.

Standard DEA does not account for multi-period optimisation.

The first two problems relate to the observed nature of the data that define the boundary of the technology and to the DEA non-parametric use of that observed data.

Parametric methods for defining the form of the transformation function use statistical mathematical models that also specify properties of the possible errors in data. Thus, *stochastic frontier analysis* (SFA) allows some observations to be above the frontier function, reducing the impact of possible outliers.

The third and fourth problems relate more to the expertise of the designer of a DEA measurement project than to DEA. The random error term in SFA can accommodate, to some extent, problems derived from omitted inputs or outputs.

For the last problem, standard DEA may be extended to account for multi-period optimisation. Chapter 4 and Chapter 5 propose the design of optimal paths of adjustment as an inter-temporal optimisation problem.

In addition to these limitations, Dyson et al. (2001, pp. 245-259) present a list of "Pitfalls and Protocols in DEA". They analyse five sources of pitfalls:

*Homogeneity assumptions* that consider three cases: non-homogeneous units, non-homogeneous environment, and economies of scale.

*The input-output set* that considers three cases: the number of inputs, outputs and firms, correlated factors, and mixing indices and quantitative measures.

*Measurement of variables* that considers four cases: percentages and other normalised data, qualitative data, undesirable inputs and outputs, and exogenous and constrained factors.

*Weights* that consider four cases: linearity assumption, zero-value weights, relative values, and linked input/output weights; and

*Weight restrictions* that consider five cases: justification of weight restrictions, non-transferability of weight restrictions, interpretation of results, absolute versus relative efficiency, and redundant weight restrictions.

## 3.7 Conclusions

In this chapter, we present the basic primal DEA model for measuring productivity. Also, we present the dual DEA model for measuring technical efficiency. Types of optimisation behaviour that are considered are profit maximisation, revenue maximisation and cost minimisation. The applicability of these concepts is illustrated with examples. DEA models are well suited for solving the problems we address in this thesis. DEA allows one to optimise objective functions of interest, subject to constraints that have implicit in them the transformation function, or the boundary of the technology. When solving problems of profit maximisation and of cost minimisation, the adjustment of inputs is not a radial reduction, but the increase of some inputs and the decrease of others may take place simultaneously.

These concepts are used in the next chapters to formulate a basic model for optimal paths of adjustment and then to extend this basic model. The basic model includes period-specific weights of firms, costs of adjustment of inputs and minimisation of the present value of costs of inputs and costs of adjustment of inputs. The extended model includes asymmetric costs of adjustment, dynamic (time-variable) outputs, costs of adjustment, input prices and output prices, and the incorporation of quasi-fixed (nondiscretionary) variables.

# CHAPTER 4

# A BASIC DEA MODEL FOR THE SELECTION OF OPTIMAL PATHS OF ADJUSTMENT

## 4.1 Introduction

Chapter 3 presents two DEA models - the primal and the dual DEA formulations. Both models measure and compare the performance of comparable firms in a set.

Section 3.2 presents the primal LP DEA model that maximises a linear mathematical expression of total factor productivity, conceptually defined as the ratio of the sum of a linear aggregation of weighted output quantities to the sum of a linear aggregation of weighted inputs quantities.

Section 3.3 presents the dual LP DEA model that optimises the technical efficiency of production. For input-orientated systems, the technical efficiency of production is defined as the ratio of the minimum quantities of inputs that are feasible to use, to the current quantities that are used; both input quantities are used to produce the same output quantity. Under this orientation, the objective is the minimisation of the quantities of input required to produce a fixed quantity of output.

For output-orientated systems, the technical efficiency of production is defined as the ratio of the current quantity of output to the maximum quantity of output that is feasible to produce; both output quantities are to be produced with the same quantities of input. Under this orientation, the objective is the maximisation of the output quantities for a fixed quantity of input. Although the DEA measurement of productivity and of technical efficiency is numerically identical they are conceptually different.

For input-orientated systems, for each non-efficient firm, there corresponds an input quantity vector that transforms the firm into an efficient one. These optimal quantities of inputs are the *input targets*. The target input quantity vector is the current input quantity vector times the technical efficiency.[1]

Under total factor productivity measurement or technical efficiency measurement, targets are achieved by radial reduction of input quantities to the boundary of best practice. The boundary of best practice is defined by observed input-output data, and defines the highest feasible technical efficiency for the set of firms under consideration.

For output-orientated systems, *output targets* are the optimal output quantities that define a firm as an efficient one. The targets are radial expansions of output quantities up to the boundary of the technology.

[1] In Section 3.3, with Figure 3.1, the special case of slacks of inputs on the boundary of technology is considered. In this case, slack reduction of inputs is after radial reduction.

When solving the problem of profit maximisation, the problem of cost minimisation, or the problem of revenue maximisation, the adjustment of inputs is not only a radial reduction but includes non-linear changes of inputs.

As mentioned in Chapters 1 and 2, the purpose of this research is to determine the optimal time-sequence of adjustments to inputs, in order to achieve the input targets for a fixed vector of output quantities. The time-sequence of adjustments is optimal in the sense that optimises an objective function of interest. An example of an objective function to be used is the present value of the net profit, considering the period-to-period revenue, the cost of inputs and the cost of adjustment of inputs.

This chapter presents a basic dual DEA model for the selection of optimal paths of adjustment. The basic dual DEA model is derived from the fundamental dual DEA problem that is presented in Chapter 3.

Section 4.2 introduces concepts used in this and following chapters: cost of adjustment, dynamic DEA, path of adjustment, period, profit, present worth value factor, 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 a dynamic DEA as a sequence of period-to-period static dual DEA problems. Period-to-period adjustment of an 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 example of the basic DEA model. Section 4.6 closes with some conclusions.

Without loss of generality, the following chapters of this thesis refer mainly to input-orientated systems. We use examples to illustrate the extension of these concepts to output-orientated systems.

## 4.2 Concepts and Definitions

This section introduces some concepts useful for the definition and the solution of the models for the problem examined in this thesis:

*Adjustment*

*Adjustment* is the variation that the input quantity vector, $\boldsymbol{x}_{kt}$, of firm *k* performs at the start of period *t*, $t = 1, 2, \ldots,$ T. The adjustment may be an increase, $\boldsymbol{x}^{+}_{kt}$, or a decrease, $\boldsymbol{x}^{-}_{kt}$, in input quantities. When it is not required to specify if the adjustment is an increase or decrease, the generic notation, $\boldsymbol{x}^{a}_{kt}$, for adjustment is used. The adjustments are the optimisation variables.

*Cost of Adjustment*

As stated in Chapter 1, in order to change the quantity of some input, such as labour, firms have to consider that to modify any input under contract can be done only by incurring internal costs to the provider and to the firm. This cost, hereafter referred to as *cost of adjustment*, is specific for each input and to each firm. The cost of adjustment of an input is different from its price. In the most general case, the cost of increasing an input is different from the cost of decreasing it. This case is referred to as *asymmetric cost of adjustment*. For example, the law tightly regulates the labour market in Chile. Compensation must be paid to any fired employee if the reason for firing is not imputable to the employee. The compensation the employer has to pay is one month of salary for every year the fired employee worked for the employer.[2] The cost of hiring a person may be high if it includes training.

This concept of cost of adjustment relates the idea of *quasi-fixed* inputs, as discussed by Treadway [1970] in the context of period-to-period variable inputs.

*Dynamic DEA*

For the purpose of this thesis, *dynamic DEA* refers to the incorporation of period-to-period variation of inputs (optimal path of adjustment), of outputs, and of technology to the basic dual DEA model. The variation of outputs and of technology may be forecasted by specialists or may be expected variations of managers. The adjustments of inputs are the result of the optimisation problem. A time sequence of standard DEA models is used to emulate dynamic DEA. Standard DEA models consider that prices, inputs and outputs do not involve time in any essential manner (Sengupta, 1995, p. 3).

*Path of Adjustment*

The *path of adjustment*, $\boldsymbol{x}_{kt}$, is the optimal sequence of input quantity vectors that a profit non-efficient firm has to develop, from the current input quantity vector up to the

[2] Recently, the time for compensation purposes was limited to 11 years, and the monthly payment to approximately AUD 2,000.

target one. The path of adjustment is optimal in the sense that it maximises the present value of the profit of the firm, discounted at a constant compound rate and over the time horizon.

*Period*

*Period* is the unit of time measurement. It is assumed that within periods all variables have constant values; nonetheless these values may change from one period to another. This assumption is generally accepted in econometric analysis. Once the span of time involved is defined, the period is irrelevant for the deduction and statement of the mathematical expressions of model. For this thesis, one period is taken to be six months, unless otherwise specified.

*Present Worth Value Factor*

Following Canada and White (1980, pp. 1, 23), because of the opportunities for investing money and increasing its value, a sum of money today is worth more than the same amount some time in the future. The timing of cash flows influences what is termed "the time value of money". To determine the rate of change of value of money the criteria is to assign the opportunity cost, i.e., the return forgone or expense incurred because the money is invested in this project rather than in other possible alternative projects. In the terminology of classical economics, the opportunity cost is a measure of the maximum benefits that, for any given situation, can be obtained from an extra unit of capital. Management assigns the rate of change of value of money as a percentage of investment by period.

For the purpose of this thesis, the cost of adjustment is considered an investment that modifies the period-to-period cash flow of the firm, and for this reason the adjustment of inputs is evaluated as an investment project.

The criterion of evaluation is to maximise the present value of total profit of the firm, along the time horizon. The factor that transforms a cash flow at some point of time to a present value is the present worth value or discount factor.

*Profit*

As previously stated, the expected objective of the firm is to maximise profit. In Chapter 2 we indicated that, for the purpose of this thesis, *net profit* is the residual after the 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 cost of adjusting inputs to their target values is included as a money cost. In this thesis, we do not include utility. The purpose is to place as few restrictions as possible on the expected behaviour of firms, so as to derive an LP formulation that is as general as possible. After target quantities have been achieved, the cost of adjustment is zero. Profit and net profit are used interchangeably.

*Time of Adjustment*

The *time of adjustment*, $t_a$, is the number of periods that a firm effectively requires to perform at the target input quantity vector. Within this time of adjustment, the initial input quantity vector, $\boldsymbol{x}_{k0}$, is transformed into the target input quantity vector, $\boldsymbol{x}_k^*$. The time of adjustment is constrained to a prefixed upper value, $t_a^*$. During the time of adjustment, the firm incurs the period-to-period costs of adjustment and takes the benefits of the period-to-period decrease of total cost of inputs.

*Time Horizon*

The *time horizon*, *T*, is the prefixed number of periods that management considers for the economic evaluation of each specific investment project. For the purpose of this thesis, the adjustments that each firm must perform to the input quantity vector are a specific investment project that is evaluated. The economic evaluation computes the present value of net profit of each project, over the time horizon.

The time horizon may be equal to or larger than the time of adjustment. If the time horizon is longer than the time of adjustment, the firm capitalises along $T\text{-}t_a$ extra periods the savings from the cost of the reduction of the inputs. In Appendix 7, we prove that a firm will perform adjustments if the present value of period-to-period reductions of total cost of inputs is larger than the present value of period-to-period costs of adjustment. The reduction of cost of inputs is evaluated with respect to the cost of the initial input quantity vector.

## 4.3 Basic Model for the Selection of Optimal Paths of Adjustment

This section presents the formulation of a basic model for the selection of optimal paths of adjustment. As indicated above, without loss of generality, the formulation is developed for input-orientated systems and a constant returns-to-scale technology. The extension to a variable returns-to-scale technology and the modification to output-orientated systems are trivial.

For input-orientated systems, given a fixed exogenous output vector, $\boldsymbol{y}_k$, the problem for firm $k$ is to choose the time sequence of input quantity adjustment vectors, $x_{kt}^{a}$, ($t$ = 1, 2, …, $t_a$), which maximises the present value of profit. This time sequence of adjustments of inputs modifies the input quantity vector, $\boldsymbol{x}_{kt}$, period-to-period from the initial input quantity vector, $\boldsymbol{x}_{k0}$, to the target input quantity vector, $\boldsymbol{x}_k^*$. The adjustments are performed within the specified longest time of adjustment, $t_a^*$. The time sequence of vectors, $\boldsymbol{x}_{kt}$, $t$= 1, 2, …, $t_a$, is the *optimal path of adjustment.*

For modelling purposes, it is assumed that all variables have constant values within each period, although they may change value from one period to the next.

For evaluation purposes, we assume that adjustment costs are incurred at the start of each period; savings from the adjustments are realised at the end of each period with the gross income from sale of products; and that inputs costs are incurred at the end of each period. Also, consistent with the assumption that values are constant within each period, we neglect transient values of variables. This means that variables behave as expected since the start of the period, i.e., any set up and start up times are negligible.

For firm $k$, at the end of any period $t$, the cost of inputs is $\boldsymbol{x}_{kt}\,\boldsymbol{w}$, where $\boldsymbol{x}_{kt}$ is the input quantity vector at period $t$, and $\boldsymbol{w}$ is the vector of constant input prices. The vector of adjustment of the input quantity vector at period $t$, $\boldsymbol{x}_{kt}^{a}$, has a total cost of adjustment of

$x^a_{kt}\ w^a$, where $w^a$ is the vector of costs of adjustment. At the end of each period, the gross income of firm $k$ is $y_k\, p$, where $y_k$ is the fixed and known output vector, and $p$ is the vector of constant output prices.

In this thesis, we assume that a firm is of a size such that it, individually, cannot influence the market of its products. As a buyer, the firm regards the price of inputs as given, for the same reason. These firms are known as *price-taking firms* (Doll and Orazem, 1984, p. 15; Chambers, 1988, pp. 50, 121).

With these revenues and costs, the profit, $\pi_{kt}$, of firm $k$ in period $t$ is:

$$\pi_{kt} = y_k p - x_{kt} w - x^a_{kt}\ w^a \qquad (4.3.1)$$

The total present value of profit of firm $k$ is:

$$\pi_k = \sum_{t=1}^{T} [s_t\,(y_k\ p\ - x_{kt}\ w) - s_{t-1}\ x^a_{kt}\ w^a\ ], \qquad (4.3.2)$$

where $s_t$ is the present value factor from period $t$.

The mathematical formulation of the problem of the firm is

$$\underset{x^a_{kt},\lambda_t}{Maximise}\ \{\pi_k = \sum_{t=1}^{T} [s_t\,(y_k\ p\ - x_{kt}\ w) - s_{t-1}\ x^a_{kt}\ w^a\ ]\} \qquad (4.3.3)$$

subject to

$$\left.\begin{array}{l} Y\lambda_t \le y_k \\ X\lambda_t \le x_{kt} \end{array}\right\} \qquad \text{(technology constraint)}$$

$$x_{kt} = x_{k,t-1} + x^a_{kt} \qquad \text{(transition equation)}$$

$$x^a_{kt}\ w^a \le b_{kt} \qquad \text{(budget constraint)}$$

$s_t = (1+r)^{-t}$ (present value factor)

$\lambda_t, x_{kt}, x_k^*, y_k \geq 0$, (non-negative variables)

where, as in Chapter 3, $X$ is the $I \times K$ matrix of input quantity vectors, and $Y$ is the $J \times K$ matrix of output quantities; $\lambda_t$ is the weight vector for the firm at period $t$; $r$ is the percentage rate of discount for present value determination; I is the number of inputs; J is the number of outputs; and K is the number of firms involved in the observations. Similar LP problems must be written for each one of the K firms.

Model (4.3.3) has $x_{kt}^a$ and $\lambda_{kt}$ as optimisation variables, and solves, in one step, the assignment of targets and the definition of the optimal path of adjustment. Hereafter, we refer to model (4.3.3) as the basic dual DEA model for optimal paths of adjustment, or just as the basic DEA model. Although the solution of problem (4.3.3) modifies, period to period, the input quantity vector of the firm under study, $x_{kt}$, the observed data that define the boundary of technology remain unmodified, including the case that the firm under examination is technically efficient.

Model (4.3.3) reduces to the profit maximisation problem (3.5.1) for T=0. With that condition, the present value of profit is the value for one period, and the input quantity vector corresponds to the target input quantity vector, and, since there are no adjustments, the cost of adjustment is zero. For this reason, we may state that the maximisation problem (3.5.1) is a particular form of model (4.3.3).

Model (4.3.3) presents two issues that deserve special attention. First, the vector of weights, $\lambda_{kt}$, is considered to change from period to period. Second, the input quantity adjustment vectors, $x_{kt}^{a}$, and the vector of weight of peers, $\lambda_{kt}$, are the optimisation variables. These are discussed in order below.

*First Issue*

The basic DEA model considers that peer(s) may change from one period of adjustment to the next. The reason for this is that technically efficient firms will crossover the firms that are on the boundary of technical efficiency, while adjusting the input quantity vector to achieve cost efficiency.

Similarly, as a technically non-efficient firm adjusts its input quantity vector to the target one, the cost of inputs, the weight of peers and the peers may change from period to period.

To illustrate this issue consider the data in Table 3.1. Suppose we wish to determine the optimal path of adjustment that maximises the present value of the profit over five periods (the time horizon) and that the adjustments to the input quantity vector are expected to be made within the same five periods.

Table 4.1 presents the original data from Table 3.1 and the costs of adjustment of inputs. Additionally, consider a rate of discount of 9.0 per cent by period and assume that the budget constrains the cost of adjustments up to \$20.0 for the first adjustment period, and up to \$30.0 for the second adjustment period.

Figure 4.1 represents the data of Table 4.1. Bold numbers 2, 1, 3 and 5 represents the inputs by unit output of firms 2, 1, 3 and 5, respectively. Assuming constant returns-to-scale technology, segments horizontal to 2, 2-1, 1-3, 3-5, and vertical from 5 define the boundary of technology. Firm 4 is a non-efficient firm.

**Table 4.1:** Inputs, Outputs, Prices, and Costs of Adjustment for Five Firms

| | Quantity at $t = 0$ | | |
|---|---|---|---|
| Firm $k$ | $y_{kt}$ | $x_{1kt}$ | $x_{2kt}$ |
| 1 | 100 | 100 | 100 |
| 2 | 110 | 90 | 149 |
| 3 | 120 | 150 | 85.9 |
| 4 | 115 | 133 | 189 |
| 5 | 103 | 152 | 161 |
| Price by unit | 8.0 | 2.0 | 3.0 |
| Adjustment Cost by unit | | 1.1 | 2.0 |

The isocost line P'P" is the same for all firms because, for this example, the price of inputs is the same for all firms. The isocost line is tangent to the boundary of technology at firm 3, determining that the input quantity vector by unit of output for firm 3 is the target input quantity vector by unit of output for the other four firms.

Figure 4.1 presents the optimal path of adjustment of firm 2. The initial inputs by unit of output for firm 2 are at vertex **2.** At these conditions, the peer of firm 2 is itself with weight 1.0. The economic efficiency at initial conditions is 0.8153.

At the start of the first adjustment period, firm 2 adjusts the inputs by unit of output from initial values **2** to $2^1$. At these conditions, the peers of firm 2 are itself with weight 0.80 and firm 1 with weight 0.22. The economic efficiency is 0.8359.

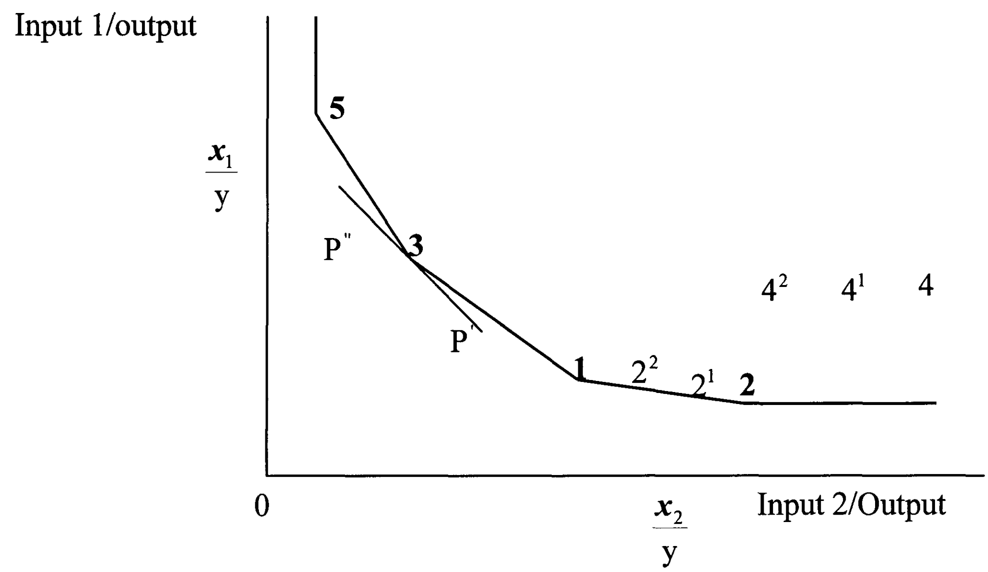


**Figure 4.1:** Optimal Paths of Adjustment for Firms 1 and 5

At the start of the second adjustment period, firm 2 adjusts the inputs by unit of output from $2^1$ to $2^2$. At these conditions, the peers of firm 2 are itself with weight 0.50 and firm 1 with weight 0.55. The economic efficiency is 0.8687.

At the start of the third adjustment period, firm 2 adjusts the inputs by unit of output from $2^2$ to **3.** At these conditions, the peer of firm 2 is firm 3 with weight 0.9167. The economic efficiency is 1.0.

From the third period of adjustment on, there are no adjustments because the input quantity vector by unit of output for firm 3 is the target one. The input quantity vectors and the peers of firm 2 are taken from the solution of problem (4.3.3) with data from Table 4.1 and economic data presented above.

Similarly, Figure 4.1 presents the optimal path of adjustment of firm 4. The initial inputs by unit of output of firm 4 are at **4.** At these conditions, the peers of firm 4 are firms 2 and 3, with weights 0.2382 and 0.7421, respectively. The economic efficiency is 0.6416.

At the start of the first adjustment period, firm 4 adjusts the inputs by unit of output from initial values 4 to $4^1$. At these conditions, the peers of firm 4 are firms 2 and 5, with weights 0.5161 and 0.5678. The economic efficiency is 0.6656.

At the start of the second adjustment period, firm 4 adjusts the inputs by unit of output from $4^1$ to $4^2$. At these conditions, the peer of firm 4 is firm 2, with weight 1.048. The economic efficiency is 0.7051.

At the start of the third adjustment period, firm 4 adjusts the inputs by unit of output from $4^2$ to **3.** At these conditions, the peer of firm 4 is firm 3, with weight 0.95833. The economic efficiency is 1.0.

From this period on, there are no adjustments because the input quantity vector by unit of output of firm 3 is the target one. The input quantity vectors and the peers of firm 4 are taken from the solution of problem (4.3.3) with data from Table 4.1 and economic data presented above.

Giving due account to the fact that adjustments involve non-radial modifications of the input quantity vector, we adopt the notation $\lambda_{kt}$ for the weight of firm $k$ at adjustment period $t$.

The conclusion is that a firm, performing adjustments, may change the peers and the weights of each peer because the adjustments to the input quantity vector may involve other than radial changes. This is a relevant aspect of this basic DEA model, because it suggests that Model (4.3.3) may be visualised as a time sequence of static fundamental DEA models that are linked by a common objective function and a sequence of transition equations.[3]

*Second Issue*

The basic DEA model (4.3.3) considers that the input quantity adjustment vectors, $x_{kt}^{a}$, and the vector of weight of peers, $\lambda_{kt}$, are the optimisation variables. With this option the input quantity vector at period $t$ is the result of decisions on adjustment of inputs, not a decision in itself. The *transition equation*, $x_{kt} = x_{k,t-1} + x_{kt}^{a}$, defines the input quantity vector at adjustment period $t$ in terms of the input quantity vector at the previous adjustment periods and the input quantity adjustment vector at adjustment period $t$. In the basic DEA model (4.3.3), the transition equations are constraints. This consideration, together with the interpretation that $\lambda_{kt}$ relates to the adjustments performed, permits us to visualise model (4.3.3) as a time sequence of fundamental

[3] The basic DEA model has a profit maximisation objective function and the boundary of technology as a constraint.

DEA models, linked by a common objective function and a sequence of transition equations.

For input-orientated systems, Figure 4.2 presents the basic dual DEA model as the mentioned sequence of fundamental dual DEA problems. The fundamental dual DEA problem is presented in model (3.5.1) and is structured as an objective function that maximises the net profit of a firm and the boundary of technology as a constraint.

The dot lines in Figure 4.2 circumscribe a fundamental dual DEA problem. The input quantity vector at $t_a$, $x_{k,t_a}$, is transformed to the fixed output quantity vector, $\boldsymbol{y}_k$.

The block, $T_0$, $\lambda_{kt_a}$, represents the DEA boundary of (static) technology, which transforms the input quantity vector to the fixed output quantity; the weight of firm $k$ is $\lambda_{kt_a}$.[4]

Input and output quantity vector, and the input quantity adjustment vector with prices and costs of adjustment, determine the net profit at adjustment period, $t_a$. The vector of decision variables (adjustment of input quantity vector), $\boldsymbol{x}_{kt}^{a}$, is generated at the time of maximising the total present value of the net profit of the firm, and modifies the *input* quantity vector from one period to the next.

[4] In Chapter 5, the constraints, *fixed output quantity vector, $y_k$,* and *static technology, $\boldsymbol{T_0}$*, are relaxed. Extensions for period-to-period changes of output quantities and technology are included.

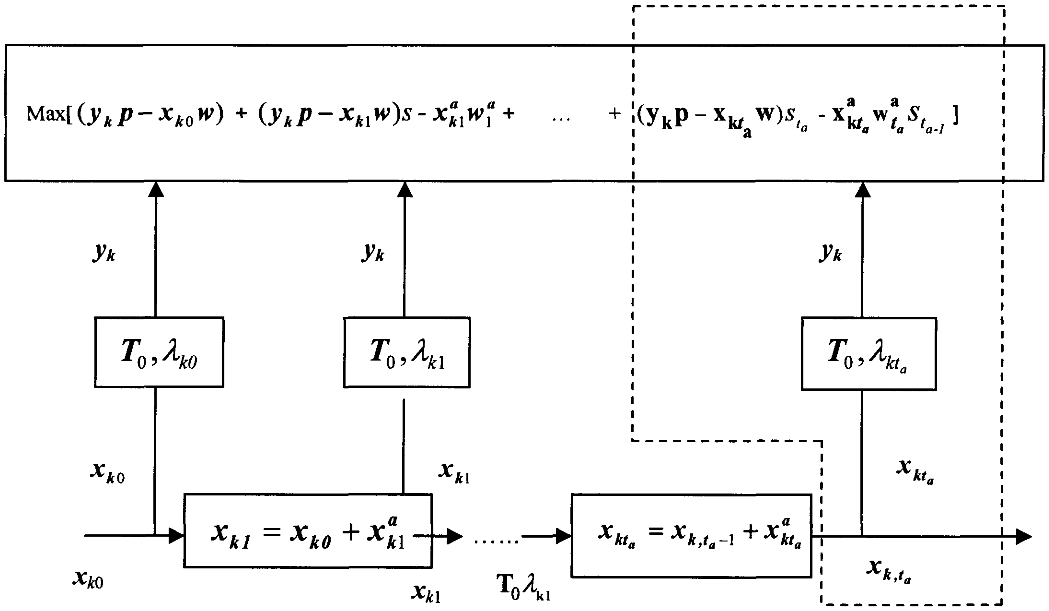


**Figure 4.2**: Model (4.3.3) as a Sequence of Fundamental Dual DEA Models

The sequence of fundamental dual DEA models is linked by the transition equations and the objective function that maximises the sum of present value of profit of each period and the cost of adjustment of inputs of each period. The input quantity vectors are constrained by both the boundary of the technology and by the optimal adjustment corresponding to the period.

## 4.4 Application of the Basic Dual DEA Model for the Selection of Optimal Paths of Adjustment

This section presents an example of the application of the basic model for the selection of optimal paths of adjustment, as is introduced in Section 4.3.

As previously mentioned, without loss of generality, constant returns-to-scale technology is assumed. The extension to variable returns to scale is trivial.

Consider the data in Table 3.1 and suppose we wish to determine the optimal path of adjustment that maximises the present value of profit over a time horizon of five periods. The adjustments to the vectors of the input quantities are expected to be made within a time of adjustment of five periods. Table 4.1 presents the adjustment costs and the original data from Table 3.1.

Consider a rate of discount of 9.0 per cent by period and assume that the budget constrains the expenditure of costs of adjustments up to \$20.0 for the first adjustment period, and up to \$30.0 for the second adjustment period.

For firm 1, problem (4.4.3) becomes:

$$\text{Maximise } \{\pi_1 = \sum_{t=1}^{5} [\, s_t\ (100 \times 8.0 - x_{11t} \times 2.0 - x_{21t} \times 3.0) \qquad (4.4.4)$$

$$- s_{t-1}(x^-_{11t} \times 1.1 + x^-_{21t} \times 2.0 + x^+_{11t} \times 1.1 + x^+_{21t} \times 2.0)]\,\}$$

subject to

the budget constraint for period 1,

$$x^+_{111} \times 1.1 + x^+_{211} \times 2.0 + x^-_{111} \times 1.1 + x^-_{211} \times 2.0 \leq 20.0;$$

the budget constraint for period 2,

$$x^+_{112} \times 1.1 + x^+_{212} \times 2.0 + x^-_{112} \times 1.1 + x^-_{212} \times 2.0 \leq 30.0;$$

the transition equations,

$x_{11t} = x_{11t-1} + x^{+}_{11t} - x^{-}_{11t}$, $t = 1, 2, \ldots, 5$,

$x_{21t} = x_{21t-1} + x^{+}_{21t} - x^{-}_{21t}$, $t = 1, 2, \ldots, 5$;

the boundary of technology, adjustment periods, $t = 1, 2, \ldots, 5$,

$100\lambda_{1t} + 90\lambda_{2t} + 150\lambda_{3t} + 133\lambda_{4t} + 152\lambda_{5t} \leq x_{11t}$

$100\lambda_{1t} + 149\lambda_{2t} + 85.9\lambda_{3t} + 189\lambda_{4t} + 61\lambda_{5t} \leq x_{21t}$

$100\lambda_{1t} + 110\lambda_{2t} + 120\lambda_{3t} + 115\lambda_{4t} + 103\lambda_{5t} \geq 100$;

the present value factors,

$s_0 = 1.0$; $\mathbf{s}_1 = (1+0.1)^{-1}$; $\mathbf{s}_2 = \mathbf{s}_1 \times \mathbf{s}_1$ ; $\mathbf{s}_3 = \mathbf{s}_1 \times \mathbf{s}_2$ ; $\mathbf{s}_4 = \mathbf{s}_1 \times \mathbf{s}_3$ ;

and

all variables are positive.

Similar LP problems must be written for firms 2, 3, 4 and 5. For computational purposes, we write, in explicit form, the increase, $x^{+}_{ikt}$, or decrease, $x^{-}_{ikt}$, of input $i$, of firm $k$ at adjustment period $t$.

Table 4.2 shows the optimal paths of adjustment for the 5 firms. In this table, we use the extended notation, $\lambda_{ekt}$, to display unambiguously the weights of peers. In the extended notation, $e$ stands for the *e*valuated firm whose profit is maximised, where $e$ = 1, 2, 3, 4, 5; $k$ stands for the peer of firm $e$, $k$= 1, 2, 3, 4, 5; and $t$ stands for the period of adjustment, $t$= 1, 2, 3, 4, 5. Note that firm 5 does not perform adjustments.

**Table 4.2:** Optimal Paths of Adjustment for the Five Firms

| | | Firm | 1 | 2 | 3[1] | 4 | 5 |
|---|---|---|---|---|---|---|---|
| | Initial Values | y | 100 | 110 | 120 | 115 | 103 |
| | Initial Values | $x_1$ | 100 | 90 | 150 | 133 | 152 |
| | Initial Values | $x_2$ | 100 | 149 | 85.9 | 189 | 61 |
| | Initial Values | TE | 1.0 | 1.0 | 1.0 | 0.7567 | 1.0 |
| | Initial Values | EE | 0.9295 | 0.8154 | 1.0 | 0.6416 | 0.9829 |
| | Target Values | $x_1$ | 125.00 | 137.50 | 150 | 143.75 | 128.75 |
| | Target Values | $x_2$ | 71.585 | 78.742 | 85.9 | 82.32 | 73.731 |
| | Target Values | TE | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| | Target Values | EE | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 1 | $x_1$ | 105.83 | 94.00 | 150 | 133.00 | 152 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 1 | $x_2$ | 93.261 | 141.2 | 85.9 | 179.00 | 61 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 1 | TE | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 1 | EE | 0.9453 | 0.8359 | 1.0 | 0.6656 | 0.9829 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 1 | $\lambda_{ek1}$ | $\lambda_{111}$=0.763 $\lambda_{131}$=0.197 | $\lambda_{211}$=0.220 $\lambda_{231}$=0.800 | $\lambda_{331}$=1.0 | $\lambda_{421}$=0.508 $\lambda_{451}$=0.575 | $\lambda_{551}$=1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 2 | $x_1$ | 114.82 | 100.00 | 150 | 133.00 | 152 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 2 | $x_2$ | 83.152 | 129.50 | 85.9 | 164.00 | 61 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 2 | TE | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 2 | EE | 0.9700 | 0.8687 | 1.0 | 0.7051 | 0.9829 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Period 2 | $\lambda_{ek2}$ | $\lambda_{112}$=0.407 $\lambda_{132}$=0.494 | $\lambda_{212}$=0.550 $\lambda_{232}$=0.500 | $\lambda_{332}$=1.0 | $\lambda_{422}$=1.048 | $\lambda_{552}$=1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Periods 3 to 5 | $x_1$ | 125.0 | 137.50 | 150 | 143.75 | 152 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Periods 3 to 5 | $x_2$ | 71.583 | 78.742 | 85.9 | 82.32 | 61 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Periods 3 to 5 | TE | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Periods 3 to 5 | EE | 1.0 | 1.0 | 1.0 | 1.0 | 0.9829 |
| OPTIMAL PATH OF ADJUSTMENT | Adjustment Periods 3 to 5 | $\lambda_{ekt}$ | $\lambda_{13t}$=0.833 | $\lambda_{23t}$=0.917 | $\lambda_{33t}$=1.0 | $\lambda_{43t}$=1.0 | $\lambda_{55t}$=1.0 |

(1) Firm 3 is peer for all firms, because it presents the largest actual profit by unit of output.

## 4.5 Some Conclusions From the Example of Section 4.4

This section presents some conclusions, derived from the solution of the above example. The solution to problem (4.3.3), for this example, suggests the following conclusions.

The solution to problem (4.3.3) gives a sequence of adjustments to inputs. The sequence of adjustments to inputs maximises the present value of profit, which is revenue minus the total cost of inputs and the cost of adjustments.

Firms performing the prescribed adjustments have specific peer(s) along each adjustment period. As stated above, the reason for the change is that while minimising the cost of inputs, technically efficient firms crossover the firms that are on the boundary of technical efficiency.

At target conditions, because prices are the same for all firms, the firms have as peers the firms that are allocatively efficient. Peer firms do not perform adjustments. In the example, the peer is firm 3. Extension to specific prices for each firm is trivial.

As stated before, there are no adjustments if the present cost of adjustments is larger than the present value of savings. This is the case for firm 5 that does not perform adjustments. In Appendix 7, we state and prove Corollary 1: A firm will perform adjustments (to input quantity vector) if the present value of savings derived from adjustments is larger than the present value of the costs of those adjustments.

While performing the adjustments to the input quantity vector the firm improves its technical, allocative and economic efficiencies. By definition of these efficiencies, once at the optimal input quantity vector, the three efficiencies are equal to one.

Table 4.2 presents the technical and economic efficiencies for the five firms, allocative efficiencies being omitted. The firms that perform adjustments, at the same time that maximises profits period to period, consistently improve their technical and economic efficiencies. Firms 1, 2, 3 and 5 are 100 per cent technically efficient for the five periods studied. This means that for firms 1 and 2 the adjustment of inputs is done over the boundary of the technology; there is no deviation into the technically feasible hyperspace. Economic efficiency measurements improve because after each adjustment the input quantity vector is closer to the target input quantity vector. The closer the input quantity vector is to the target input quantity vector, the lower is the distance between the input quantity vector and the minimum cost isocost plane of that firm. This plane is tangent to the boundary of the technology at the peer of the firm involved.

Firm 4 initially is technically not efficient. After the first adjustment period, the technical efficiency measurement improves from 75.67 to 77.39 per cent; at the same time, the economic efficiency improves from 64.16 to 66.56 per cent. After the second adjustment, this firm has 100 per cent technical and economic efficiencies.

## 4.6 Conclusions

The basic DEA model solves, in one step, the design of optimal paths of adjustment. The basic DEA model considers that peer(s) change from one period to the next, the reason for this is that technically efficient firms crossover the firms that are on the boundary of the technology while increasing to cost efficiency on the boundary of the technology. Changes include weight values of peers, deletion of a peer and the incorporation of a peer. This is a relevant aspect of this basic DEA model, because the design of optimal paths of adjustment may be considered as a sequence of basic DEA models linked by the transition constraints and one common objective function. On the consideration that the basic DEA model is a sequence of basic DEA models, it is possible to modify each basic model of the sequence as required. Modifications of interest are period-to-period changes of the output vector and period-to-period changes of the technology. Chapter 5 extends the basic dual DEA model with these modifications.