OXFORD INSTITUTE FOR ENERGY STUDIES

Evaluation of Improvements in Petroleum Refining Technology: Incentives, Research & Development and Outcomes

J. L. Enos

Oxford Institute for Energy Studies

EE18

1994

The contents of this paper
are the author's sole responsibility.
They do not necessarily represent the
views of the Oxford Institute for
Energy Studies or any of its Members.

ISBN 0 948061 81 2

CONTENTS

FIGURES

ABSTRACT

The Fluid Catalytic Cracking (FCC) process was introduced commercially in 1942; in the half-century of its existence, it has been improved again and again, so that plants today operate on a much larger scale and at a fraction of their original unit costs. This paper provides a *methodology* for a study of the history of improvements to the FCC process, to understand both their engineering and economic natures.[1]

Improvements in an industrial process can take many forms – new pieces of equipment, greater knowledge of scientific and engineering principles, more efficient organization and administration of resources, and so on. To relate these forms to each other is a difficult undertaking, yet necessary if the overall sequence is to be understood. In this paper the problem of relating different improvements in their sequence is discussed; and a common mode of representation is suggested, involving the application of linear programming. It is argued that the shadow prices of the constraints in the dual programme give indications of the incentives for improvements, and that the changes in the coefficients of the technical matrix give indications of their consequences. For an understanding of the creation of the improvements themselves, however, one must still rely upon a history of the industry's research and development activities.

Two problems in applying linear programming to the analysis of a changing economic system are discussed. The first arises in expressing the non-linear technical and economic relationships in linear form; the second in representing as a single system what is, because of the effects of the improvements, an evolving one. These problems, and the ways in which they will be addressed, are illustrated at the end of the paper.

Finally, a method of planning companies' Research and Development activities frequently applied in the industry is reported, and its informational requirements identified. The extent to which the data generated in the course of this study of improvements will be useful in such exercises is then indicated.

[1] Other parts of the study will be published in subsquent papers and the final work will appear as a book.

1 INTRODUCTION

In 1942, the Fluid Catalytic Cracking process was presented to the petroleum refining industry. At the time of the innovation, it was no more profitable to operate than existing processes; it is only in the fifty years between then and now that its great potential has been realized. Realization has come through a stream of improvements, not one of which has equalled the genius of the concept of the fluidized bed, but all of which have required acute awareness and substantial effort.

Great novelty evokes fascination for the observer; minor variations beget boredom. This is true of fashion; it is also true of studies of technical change. The great novelties, the innovations, attract the attention of historians and economists, whereas the successive improvements to those innovations are generally neglected. To be sure, it is recognized how important improvements are in their cumulative effects (Rosenberg, 1982; Freeman, 1982). Yet systematic studies of a sequence of improvements are few and far between; the earliest and most detailed was that of Hollander (Hollander, 1965); more recent ones have examined improvements in the manufacture and sale of computers and in the operation of nuclear power plants (Noyce, 1977; Lester and McCabe, 1993).

However impressive the improvements, and however significant they may be in reducing costs and raising quality, the mechanism by which they are brought about remains undiscovered. The technical changes underlying the improvements are usually described in narrative form and their scientific and technological aspects are emphasized; the economic consequences are illustrated by cost calculations, comparing costs before and after the improvements were instituted. The association between the technical nature of the improvements and their economic incentives and consequences remains vague.

In this study of improvements in the Fluid Catalytic Cracking process throughout its half-century of operation, the objective is to relate the technical changes with the economic. Put one way, the aim is to insert the technology into the economic environment; put another, the aim is to determine the economic dimensions of the change in technology. It is hoped that this will provide economists with a better understanding of the nature of the technologies with which they contend, and engineers with a better understanding of the economic forces that impel technological developments.

The unit of enquiry for the study of improvements in petroleum refining technology is the process. By process is meant a technology plus its application; i.e. the physical and chemical relationships governing the transmutation of elements, the pieces of equipment and other inputs in which these transmutations take place, and the sum of the knowledge that is necessary in order to organize and conduct the activities. The process may be very simple, incorporating no more than a single unit operation (to use the chemical engineer's term); it may be extraordinarily complex, employing many, inter-related unit operations and processes in expensive equipment operating at a very large scale, consuming a variety of inputs and producing a variety of outputs, carefully controlled and efficiently administered. In the petroleum refining industry, the process is delineated in such a way as to enable it to be carried out in a cluster of pieces of equipment on a single site and operated by a team of highly trained individuals. The word 'process' thus has several meanings: it can refer to the technology that is employed; it can encompass all the pieces of equipment collected on a single site; it can signify the group of individuals who have the responsibility of operating the equipment. It may occupy an entire industrial plant, or it may be a portion of the plant, where it is combined with other processes. It may change through time, employing different physical or chemical relationships, utilizing different pieces of equipment, being organized in different ways; but the changes are not usually of such great significance that the process loses any of its major characteristics.

2 ECONOMIC AND ENGINEERING PRODUCTION FUNCTIONS

To discover how a petroleum refining process changes over time, in both its technological and its economic dimensions, is an extremely complicated task. To generalize over several changes is even more difficult. Although something may be lost, or errors may be made, the task is made a little easier by simplification; but it is imperative to explain the nature of the simplification. We shall try to reveal this by means of a formalization, employing the concept of the production function.

Economists have long used the concept of the production function, expressing the relationship between the inputs to, and the outputs from, a process. If the inputs are labelled as **x** and the outputs as **y**, and both are expressed in physical terms, the production function can be written as

$$f(\mathbf{x};\ \mathbf{y}) = 0 \qquad (1)$$

where **x** = vector of quantities of inputs, of dimension m,

y = vector of quantities of outputs, of dimension n, and

f = the function relating inputs to outputs.

So as to be able to obtain analytic solutions to models that include production functions, economists always assume that the function f has simple mathematical properties. Yet, the function f is a compendium: it contains all that is relevant to the process, both technological and economic. It is therefore too abstract for our purposes; we need a more complex expression that distinguishes the technological from the economic. Such a distinction can be made by decomposing equation (1), in the manner in which a petroleum refiner does when planning his production. First, he estimates the capabilities of his plant, given the various inputs that he is able to command. Let these be designated by l of the total of m types of inputs; i.e. inputs 1, 2 ... l are necessary to finance, construct and operate the plant in which the process is carried out. We can express the relation between l of the inputs **x** and the plant's capabilities **v** via the function g, where g, like the preceding function f, is a transformation from what is applied (the **x**) into what results (the **v**).

$$g(\mathbf{x};\ \mathbf{v}) = 0 \tag{2}$$

where $\mathbf{x}$ = the first l components of the vector of inputs,

$l \leq m$,

$\mathbf{v}$ = vector of process capabilities, of dimension k, and

g = the function relating the first l inputs of $\mathbf{x}$ to the capabilities.

If we keep our variables in terms of rates or flows, the inputs $\mathbf{x}$ will be so many man-hours, so many units of land rental, so many units of capital service, etc. What of the capabilities? Here it may be useful to choose a concrete example, albeit the simplest possible one. Imagine that the process under investigation is the pumping of a fluid (crude oil) through a pipeline (Cookenboo, 1955). For this process, capabilities $\mathbf{v}$ can be measured along just two dimensions, that of pipeline diameter and that of horsepower. (The larger the diameter of the pipe through which the oil flows and the greater the horsepower that impels the oil, the greater the capability of the pipeline. Equation (2), expressing the capabilities of the pipeline, would thus have many inputs $\mathbf{x}$ (numbering 1, 2..;) and two capabilities v_1 (line diameter) and v_2 (horsepower)). Chenery, who first made the decomposition, called equation (2) the engineering production function (Chenery, 1949).

The petroleum refiner is not content knowing only the engineering production function; he also needs to know how to exploit his plant's capabilities. In symbolic form this knowledge is expressed as

$$h(\mathbf{x};\ \mathbf{v};\ \mathbf{y}) = 0 \tag{3}$$

where $\mathbf{x}$ = the remaining elements of the vector of inputs

$(l + 1, l + 2 \ldots m)$,

$\mathbf{v}$ and $\mathbf{y}$ as before, and

h = the function relating $\mathbf{x}$, $\mathbf{v}$, and $\mathbf{y}$.

If we wished to and if equations (2) and (3) had the proper mathematical properties, we could combine equations (2) and (3), yielding equation (4),

$$f(\mathbf{x};\ \mathbf{y}) = g(\mathbf{x};\ \mathbf{v})h(\mathbf{x};\mathbf{v};\mathbf{y}) = 0 \tag{4}$$

but this is not our purpose. Rather we wish to examine more carefully equations (2) and (3), since these are abstractions from the (much more complicated) calculations of the refiner.

Returning to our objective in carrying out the study, it is to analyse improvements in petroleum refining technology, in order better to understand their nature. There is also the hope that some pattern will emerge, from what seems, on the surface, to be a *pot-pourri* of accomplishments. How might decomposing the economist's production function into the two sub-functions (the engineering function of equation (2) and the economic function of equation (3)) help in meeting the task?

There are two ways in which the decomposition helps to analyse improvements, the first because it models the refiner's efforts at improvement, the second because it provides a useful structural division.

Considering the petroleum refiner's planning, he customarily isolates those activities leading to improvements in his technology from those leading to improvements in operation; the former being called R & D, the latter 'technical service'. The former activities take place in a research laboratory, sometimes located in, more often geographically separate from, the refinery itself; the latter takes place within the refinery offices or on the process units. The former are carried out chiefly by scientists, or scientists-cum-engineers; the latter by process engineers and operators. Administratively, in large firms, they are likely to be placed in different parts of the hierarchy, the laboratory in a staff division reporting to a vice-president of R & D; the technical service group in line with the refinery manager.

Besides an organizational distinction between the different activities represented by the engineering and economic production functions, there is also a distinction between the types of improvements that are a consequence of these activities. The distinction is by no means sharp, but generally shifts in the engineering production function are brought about by improvements in process design and equipment (for example, a more efficient pump, a less-easily eroded refractory), whereas shifts in the economic production function are brought about by improvements in process operation (a higher temperature or pressure, a more nearly perfect synchronization of unit processes or operations, and so on). Sources of information on improvements will tend to be separated too; evidence on shifts in the engineering production function will be found in the refiners' R & D divisions and in the accomplishments of equipment suppliers; evidence of shifts in the economic production function will be found in refinery plant records.

3 SIMPLIFYING THE PRODUCTION FUNCTIONS

The production functions of equations (2) and (3) are implicit: if we are to evaluate improvements we must make them explicit. Where petroleum refining processes are concerned, to do so with any great accuracy has been, with one exception, impossible. (Refiners are currently devoting much energy to formulating engineering production functions based on the physical and chemical reactions carried out in distillation, cracking, etc.; but so complicated are these reactions that even representative or token reactions of clusters of individual hydrocarbons exceed the knowledge of the scientists and the computing ability of large-scale electronic computers (Maples, 1993)).

For one of the simplest of all processes (the transport of crude oil in a pipeline), for which the engineering production function can be derived from physical principles, there has been an attempt to represent the effects of improvements (Pearl and Enos, 1975); but, so far as the author is aware, this is the only application. Moreover, the application is for a host of improvements, made over a span of seventeen years, rather than for a sequence of improvements, identified separately.

The exception to the statement that petroleum refining processes are too complex to be represented by explicit functional forms is the blending of motor fuels. It appears that the blending of different components can be expressed in a linear, separable form. Since no chemical reactions take place, and since the blending components are all liquids at atmospheric temperature and pressure, the volume of any number of components, blended, is equal to the sum of the volumes of all the individual constituents:

$$x_1 + x_2 + \ldots x_b = y \tag{5}$$

where the x_i are the volumes of the constituents and y the volume of the blend.

Similarly, and conveniently, a characteristic of the blend (e.g., motor octane number) can be quite conveniently estimated as the arithmetically weighted sum of the characteristic of each component; i.e.

$$P_1 \, r^j_{x_1} + P_2 \, r^j_{x_2} \ldots P_b \, r^j_{x_b} = r_y^j \tag{6}$$

where P_i = proportion of the ith component in the blend,

$$i = 1, 2 \ldots b; \sum_{1}^{b} {}_i = 1$$

$r^j_{x_i}$ = the value of the characteristic j for the ith component, x_i, and

r^j_y = the value of the characteristic for the blend.

Both equations (5) and (6) can be written in vector form:

$$\mathbf{I'x} = y \tag{5a}$$

where **I** is the unit vector, and

$$\mathbf{P'}\ \mathbf{r^j_x} = r^j_y \tag{6a}$$

Moreover, if there are more than one characteristic, they can be considered separately. Again, with reasonable accuracy, the system of all characteristics (e) can be represented by a set of simultaneous linear equations, written in vector form as:

$$\mathbf{P'R_x} = \mathbf{r_y} \tag{7}$$

where **P** = a vector of proportions of the ith component in the blend, of dimension b,

$\mathbf{R_x}$ = a (b x e) mixture of individual characteristics,

$\mathbf{r_y}$ = a vector of characteristics of the blend, of dimension e.

The linear separable nature of the system of equations (7) was well known in the petroleum refining industry, so it is not surprising that gasoline blending was the first application of the powerful mathematical technique of linear programming. Both in universities (Charnes, Cooper and Mellon, 1952) and in the industry itself (Symonds, 1955) the blending of motor fuels was formulated as a linear programming problem; and textbooks on linear programming universally provided it as a vivid illustration of the technique (e.g. Hadley, 1962). Not surprisingly, attempts were made to formulate other scheduling problems faced by petroleum refiners as linear programmes (Symonds, 1955); in fact it was seen that an entire petroleum refinery's schedule could be so formulated (Manne, 1956; Adams and Griffin, 1972).

Linear programming recommends itself to us also, both because it has been generally applied in the petroleum refining industry for forty years, and because it will enable us to estimate incentives to improvements, and consequences thereof, at various points in that forty year interval. Let us therefore lay out the model and observe its usefulness.

4 THE LINEAR PROGRAMME AND ITS DUAL

The conventional way of writing the primal problem of linear programming is

$$\text{maximize } z = \mathbf{cx}$$

subject to

$$\mathbf{Ax} \leq \mathbf{d}$$

and

$$\mathbf{x} \geq 0 \tag{8}$$

where z = the value of the objective function (a scalar)

c = the unit value of each primal variable (a vector, usually of prices)

x = the primal variables (a vector of dimension j)

A = the matrix of constraint coefficients a_{ij} (the matrix has i rows and j columns)

d = the 'capacities' of the constraints (a vector of i components), and

0 = the null vector.

The dual to this linear programming problem is

$$\text{minimize } Z = \mathbf{d'w}$$

subject to

$$\mathbf{A'}\ \mathbf{w} \geq \mathbf{c'} \tag{9}$$

where Z = the value of the dual (a scalar)

d', **A'** and **c'** = the transposes of **d**, **A** and **c** respectively,

and **w** = the dual variables (a vector, commonly conceived of as shadow prices)

(The nomenclature is taken from Hadley, 1962, pp. 222 ff.)

If one of the pair of problems equations (8) and (9) has an optimal solution, the other also has an optimal solution. Let $\hat{\mathbf{x}}$ stand for a feasible solution to equation (8) and $\hat{\mathbf{w}}$ a feasible solution to equation (9); if

$$\mathbf{c}\hat{\mathbf{x}} = \mathbf{d}'\hat{\mathbf{w}} \tag{10}$$

then $\hat{\mathbf{x}}$ is the optimal solution to equation (8) and $\hat{\mathbf{w}}$ to equation (9). The values of the objective functions z and Z are thus equal, i.e., the maximum of the primal problem is equal to the minimum of the dual problem.

In the optima, certain properties hold, one of which we shall draw upon. The property is that of complementary slackness (*idem*, pp. 239–241). Imagine that we convert the first system of inequalities $\mathbf{Ax} \leq \mathbf{d}$ in equation (8) to equalities by adding s 'slack' variables to the r variables comprising the vector $\mathbf{x}$; then at the optimum

$$\mathbf{w}'\mathbf{x}_s \equiv \sum_{i=1}^{s} w_i \, x_{r+i} = 0, \quad i = 1, 2, \ldots s \tag{11}$$

In equation (11), if any slack variable is different from zero, its shadow price (the component w_i) will be zero; if the slack variable is zero (i.e., if there is no slack), the shadow price will be strictly positive. The analogous condition holds for any slack variable added to convert the constraints in equation (9), the dual problem, to equalities; when this is done the following condition also holds at the optimum:

$$\mathbf{x}'\mathbf{w}_s \equiv \sum_{j=1}^{s} x_j \, w_{s-j} = 0 \tag{12}$$

The uses to which equations 10–12 will be put will be explained later.

5 FORMULATING COMPARABLE EQUATION SYSTEMS

If we are to evaluate the effects of successive improvements by appeal to the technique of linear programming, we must have a single system capable of reflecting the potential of each technology. Thus it must, using the same variables, be able to describe accurately the first technology (the innovation), the second (the first improvement), the third (the second improvement) and so on, through the last improvement. Yet it must also be the same system, in the sense that it contains the same variables and is expressed as the same 'problem'.

We can clarify these conditions by referring to the linear programming model of the previous section. When we define the vectors **x**, **w**, **c** and **d**, and the matrix **A**, i.e., when we attach meanings to each one, we have specified the system: the matrix **A** specifies the technology; the vector **d** specifies the limits that physical laws, human abilities and social mores impose on the application of the technology; the vector **c** the economic conditions (prices, etc.) governing the operation of the process; and the vectors **x** and **w** specify, respectively, the quantities and values of the scarce inputs and outputs.

In principle, **A**, **d**, **c**, **x** and **w** can vary, and in practice all do. We could say that variations in **A** signify technical change, in **d** changes in scale, in **c** changes in market conditions and in **x** and **w** changes in the composition of inputs and outputs, in response to previous changes in **A**, **d**, and **c**. But this would be an oversimplification, in two senses. It is an oversimplification in the sense that changes in one element, say the technological matrix **A**, are not independent of changes in another element, say the vector **d**. It is also an oversimplification in the sense that the causality implied by saying that **x** and **w** respond to changes in the other elements may be reversed; for example, market conditions – especially the vector **c** – may change as a consequence of changes in the quantities and values taken on by **x** and **w**. In the long run, it may well be changes in the matrix **A** that drive the system.

Since we are primarily interested in technical change we shall neglect the long run; we shall eschew technological determinism, and assume that changes in economic conditions – in **c** – are exogenous to our system of equations (8) and (9). To the extent that prices do alter historically, we shall acknowledge the changes, but we shall not attempt any explanation: changes in economic conditions will be assumed to occur outside our system.

The changes that we will attempt to explain are thus the changes in **A**, **d**, **x** and **w**, singly and jointly. But first we have to determine what have been the changes in these variables and what has caused them. Our procedure will not follow this two-step sequence, however, but will take three steps: the steps will involve (1) the recording of inducements, (2) the determination of consequences, and (3) the securing of the improvements. The first and second steps will be made with the aid of linear programmes; the third chiefly with the history of events.

6 INDUCEMENTS TO IMPROVEMENTS

Considering first inducements to improvement: these are recorded by the shadow prices (**w**) of the constraint variables in the primal problem of the linear programme. At any instant, t_o, the optimal solution to the linear programme will yield a set of quantities $\hat{\mathbf{x}}_o$, which reflect the most profitable allocation of the scarce resources in the system (the **d**). From the solution to the dual programme, the shadow prices or opportunity costs of the constraints (the $\hat{\mathbf{w}}_o$) are simultaneously derived; these reflect the increases in the value of the objective function of the primal problem attributable to a unit increase in the capacities of the constraints. If the constraint is binding in the optimal solution, its opportunity cost will be strictly positive; if not, it will be zero, both according to equation 12. The magnitude of the opportunity cost is a measure of the incentive to loosen the bind: the higher the value, the greater the incentive.

Shadow prices, or opportunity costs, are not the whole measure of incentives, for they only indicate the *unit* value of relaxing a binding constraint. One also needs to know the number of units by which the particular constraint, say d_i, can be relaxed before another constraint, say d_j, begins to limit still further increases in the objective function. In a perfectly designed and operated process, i.e. one perfectly synchronized, *all* the constraints would be binding in the optimal solution, and a relaxation of d_i of one unit would not increase the value of the objective function unless the other d_js were relaxed too. A multiplicity of improvements (or one blanket improvement) would be necessary to increase profitability.

In the absence of perfect synchronization, an increase of Δd_i might be possible before another constraint, d_k, became binding. The combined measure of the incentive to secure the improvement in the activity i would then be $(\hat{w}_i)(\Delta d_i)$, and this could be compared with an estimate of the cost of securing the improvement, in order to appraise its attractiveness.

If the system of constraints is complete, reflecting not only the physical and chemical requirements of the process but also its legal, organizational, environmental and financial requirements, the $(\hat{w}_i)(\Delta d_i)$ would be comprehensive, providing all the data needed to plan improvements. But such completeness is not possible; in practice, the system of constraints reflects the engineering relationships only, and those only as understood at the time (t_o). This problem of the comprehensiveness of the linear programming model limits the types of incentives that can be identified.

7 THE CONSEQUENCES OF IMPROVEMENTS

In principle, our third concern, the consequences of any improvement, are measured by the changes in the process and its operation; these in turn are measured by the changes in the linear programme and its optimal solution. If equations (8) and (9) of section 4 describe the process at time zero, before the improvement is made, a new linear programme will describe it at time t_1, after the improvement has been made. The differences in the linear programmes are twofold: the matrix of technical coefficients **A** can have changed to $\bar{\mathbf{A}}$ and the vector of constraints **d**, to $\bar{\mathbf{d}}$.

$$\max\ z = \mathbf{cx}$$

s.t.

$$\bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{d}}$$

and

$$\mathbf{x} \geq \mathbf{o} \tag{13}$$

and

$$Z = \bar{\mathbf{d}}'\mathbf{w}$$

s.t.

$$\bar{\mathbf{A}}'\mathbf{w} \geq \mathbf{c}' \tag{14}$$

The differences between equations (8) and (9), on the one hand, and of (13) and (14) on the other, represent a menu of possible changes (given that the vector of prices **c** is invariant and the vector of input and output types **x** is comprehensive) but all such changes may not occur with a single improvement. It is conceivable that only some elements of **d** change (to $\bar{\mathbf{d}}$); this could be the consequence of the recognition of the existence of a 'bottleneck', and its subsequent removal. The improvement would have been diagnostic – the application of known methods of chemical or physical analysis, or the development of a new method of analysis.

It is also conceivable that only the elements of the matrix of technical coefficients **A** change (to $\bar{\mathbf{A}}$). Changes could occur through a better understanding of the nature of the existing technique or through the successful research and development of a new mode of operation. Finally, and most probably, both the matrix of technical coefficients **A** and the vector of capacity constraints **d** could change simultaneously.

Economists have singled out particular changes and given them names: an equi-proportionate reduction in all the a_{ij} in the **A** matrix is known as Hicks-neutral technical change; equi-proportional reductions in those constraint equations reflecting the availability of capital, and in those reflecting the availability of labour, are known as Harrod- and Solow-neutral technical change respectively. Increases in one or more elements of the vector of capacity constants (again perhaps through a better understanding of the nature of the process; or through an expansion of items of capital equipment, into a novel and greater range of operation) **d** are scale improvements; and equi-proportional increases indicate constant returns to scale.

The three neutrality types are polar cases; we would be surprised to discover that any single improvement, let alone a sequence of improvements, conformed to one case or the other; although discovering from the evidence a tendency towards a single type would be interesting. Certainly we do expect to find changes in the vector **d**, particularly for specific plants operated over an extended interval.

If we wished to include all the effects of improvements in a single measure it would be in the technical matrix **A**. To consolidate the improvements in **A**, we would have to absorb within it the changes in the vector of constraints **d**; this could be achieved by normalizing **d**, and representing all improvements, whether of scale or of other types, by appropriate reductions in the elements a_{ij} of **A**. In practice, this would be equivalent to assuming a plant fixed in its physical dimensions (for a typical petroleum refinery process, say, this would signify the original major vessels in their original lay-out).

8 THE SECURING OF IMPROVEMENTS

Incentives and outcomes can be represented by successive linear programmes, but securing improvements almost always, if not always, requires an investment of resources: R & D; or the purchase and installation of improved equipment; or both. These investments take place over an interval, as do the returns that justify them. Only a dynamic model will represent adequately the commitment of resources to the securing of improvements, and the returns to these commitments.

Linear programming, in the form of a sequence of programmes each indicated by the interval of time to which it relates, can be used *ex post*, to measure the profitability of a commitment *after* it has been made and the returns obtained. In such cases, the primal problem becomes

$$z = c^1 x^1 + c^2 x^2 + c^3 x^3 \; . \; . \; . + c^T x^T$$

$$\begin{array}{c} \mathbf{A}^1 \mathbf{x}^1 \le \mathbf{d}^1 \\ \mathbf{A}^2 \mathbf{x}^2 \le \mathbf{d}^2 \\ \mathbf{A}^3 \mathbf{x}^3 \le \mathbf{d}^3 \\ \vdots \\ \mathbf{A}^T \mathbf{x}^T \le \mathbf{d}^T \end{array}$$

and

$$\begin{array}{c} \mathbf{x}^1 \ge 0 \\ \mathbf{x}^2 \ge 0 \\ \mathbf{x}^3 \ge 0 \\ \vdots \\ \mathbf{x}^T \ge 0 \end{array} \qquad (15)$$

where the dimensions of the variables **c**, **x**, **A** and **d** are now greater than those in equation (8), containing in addition the resources allocated to R&D, or to the buying of new equipment, or both. (Discounting future returns could easily be accommodated in the variables $\mathbf{c}^t$ in the form of an expression such as $\mathbf{c}^t = \mathbf{c}^1/(1+\delta)^t$ where δ would be the discount rate.)

As equation (15), the linear programme would just be encapsulating history; it would merely be a description, in the abstract, of past events. It would not be exploiting the power of linear programming to select the best from among alternative allocations or commitments. It would measure the outcome, given the decisions that led to the outcome; but it would not give any indication of the other possible decisions, some, perhaps, superior, that might have been made. It could not represent a 'make-or-buy' decision (in this case to undertake the R&D oneself, or to let some other firm conduct the R&D and then license the improvement); it could not represent the alternatives between a normal course of R&D and a 'crash programme' (in this case to commit the resources to R&D over the normal interval, or to commit a larger volume of resources – more scientists, more engineers, more laboratory space, more working capital – within a shorter interval).

If, as the administrators of R&D do, we had to evaluate alternatives such as those illustrated above, and we wanted to calculate the optimal course of action (i.e., if we wanted to rewrite the objective function of equation (15) as maximize $z = ...$, and solve the resulting system of equations) we would be destroying the linearity of the system and attempting the solution of a mixed integer programming problem, a much more demanding task. But since we are most unlikely, in our research, to be able to find data reflecting those alternatives <u>not</u> selected, this problem need not concern us. As it is, we shall find it difficult enough to encapsulate history, to reproduce past events.

9 PROBLEMS IN THE APPLICATION OF LINEAR PROGRAMMING

In the course of formulating, solving and interpreting successive linear programmes we shall encounter two major problems, for neither of which will we have completely adequate solutions. In order, these are problems of non-linearity and time irreversibility.

With the exception of the few simple operations of blending of motor gasoline, diesel fuel and heavy fuel oil, all the physical and chemical processes utilized in petroleum refining are non-linear in the variables involved. Consider the unit operations and processes covered in any chemical engineering handbook or petroleum refining textbook (Nelson, 1958); although much simplified, their mathematical formulae are still highly non-linear in form, even after transformation (say, into logarithms). Textbooks in linear programming generally avoid the issue of non-linearity, assuming for convenience that the expressions in both the objective and constraint equations are linear to begin with. (The exception is the text by Chvatal (1983), who devotes a chapter to linearization.)

If we are to take advantage of the properties of the optimal solutions to linear programmes, we must therefore linearize the formulae representing the physical and chemical operations carried out. There are two possible approaches, both of which we will follow: the historical and the synthetic. So useful has linear programming been to petroleum refiners that most have applied the technique to scheduling their production; a few have even permitted their programmes to be published. We will reproduce these from the historical record wherever possible, associating the linear programming model with the technology employed by the refiner and assuming that the model has been most accurately constructed – in particular that the model's constraint equations are the best linear representations of the (non-linear) physical and chemical relationships.

The other approach that we will use, in the absence of a published linear programme appropriate for the technology to be investigated, is the synthetic. In brief, we will construct our own programme, compare its optimal solution with best commercial practice and, assuming the latter to be the real optimum, correct the synthetic programme (meaning the system of constraint equations) so that its optimum coincides with the refiner's practice. At this point, where the primal and the dual programmes have their optima, the incentives for improvements are indicated by the shadow prices of the (binding) constraints: these are, within the neighbourhood of the optima, assumed to be linear. Mathematically, the optima are at vertices of the two convex hulls (the primal and the dual), and the incentives are the gradients of the supporting hyperplane.

Linearization thus takes place in two stages, first in the formulation of the constraint system and second in the estimation of the shadow prices: the first stage may involve considerable statistical manipulation, the second requires only the assumption of linearity.

The linear programmes developed and applied by refiners differ, depending upon the extent of their knowledge. Historically, knowledge of refining processes has increased with time, and the dimensions of linear programmes have increased *pari passu*. In terms of the mathematical equations (8) and (9), the increase in technical knowledge is expressed by ever-larger matrices **A**. The increase in the number of rows of **A** reflects greater understanding of the nature of the physical and chemical operations and of the limits that equipment imposes in their fulfilment; the increase in the number of columns of **A** reflects a greater awareness of the heterogeneity of the hydrocarbons processed in the refinery and of their specific properties.

But the historical increase in the size of the matrix of technical coefficients **A** creates the second problem in applying linear programming, that of the irreversibility of time. The problem does not arise in measuring incentives to improvements, since these are estimated with the knowledge available at the time of the estimate, and the knowledge is reflected in the contemporaneous linear programme. The problem arises in measuring the implications of improvements, and is the more serious the longer is the span of time over which improvements are surveyed. In principle, the implications of the improvements are measured by the changes in the coefficients a_{ij} of the constant matrix **A**, and, for a single plant, changes in the constraint vector **d**. Yet **A** changes with knowledge of the technology; the matrix **A** at the time of the original innovation is very different in dimension from the matrix **A** after a sequence of improvements. Presumably, the terminal matrix **A** and vector **d** will be the most nearly complete renditions of the improved technology; to account for all the improvements made up to that last instant, a matrix of the same dimension, with the same number of a_{ij}, should be available for the first, unimproved technology. But such an expansive matrix, incorporating all the accumulated knowledge, was not then available, because the knowledge had not been accumulated.

The only way to measure all the improvements, over the entire span of investigation, is to formulate a linear programme covering the initial technology (and plant) on the presumption that all future knowledge was available at the beginning. Having done this, we can then compare the a_{ij} at beginning and end, and observe the nature of the changes; i.e., the implications of the sequence of improvements. But we must remember that we have seen the implications by hindsight, and that they are not necessarily what were foreseen, nor what were sought. We cannot reverse time for our own convenience.

10 CONCLUSION

To conclude, let us give an indication of what the evaluation of a small improvement might resemble, and how it would fit into the entire sequence.

Imagine that we are at time t(k), sometime after the innovation. Imagine also that there exists at t(k) a linear programme and its dual describing the current technology, expressed in the form of equations (8) and (9). Then, according to equation (12), there would be (at least two) inducements to improvements, forecasts of perspective benefits. These are forecasts, made before the improvements emerge, and thus made *ex ante*.

More or less simultaneously, *ex ante* measures of the costs of obtaining the improvements will be made by some refiners, or process design and construction firms, or equipment suppliers, or two, or all three. Comparisons of *ex ante* benefits and costs will be made, and, where the former exceed the latter and the likelihood of success is high, R & D will be undertaken, extending over several periods.

Once the improvement emerges, say at t(k+m), it may be incorporated in the engineering process; if it is, its profitability can be estimated, for the first time, *ex post*. Successive values of the differences between benefits and costs taken at regular intervals t(k+m+1), t(k+m+2), etc., will provide even more accurate measures of the profitability of the improvement.

From the point of view of anyone wishing to evaluate improvements, there are several possible comparisons, all of which would be useful but none of which would alone be satisfactory. One could make pairwise comparisons of *ex ante* estimates (*ex ante* benefits vs. *ex ante* costs, at t(k)), or of *ex post* estimates (*ex post* benefits vs. *ex post* costs, at any date t(k+m+p), $O \leq p \leq \infty$), (see Figure 1). One could compare *ex ante* and *ex post* estimates of benefits, or of costs, or of the difference between them, (profitability). One could expand the number of estimates, so as to take in benefits and costs, *ex ante* and *ex post*, in a three- or four-way comparison.

Let us conclude our illustration by conceiving of a four-way comparison, at an intermediate time after the original innovation (and indicated by dots surrounding the letter S in Figure 1).

On the second drawing, Figure 2, the point at which the comparisons are to be made is labelled t(k). The cumulative effects of past improvements, for which there are, at t(k), *ex post* measures, are indicated by small crosses, commencing at the date of the innovation, t(o). The solid curve, up to t(k), is derived by fitting the observations, indicated by the crosses, econometrically; and the extension of the solid curve beyond t(k) is an extrapolation, based on the hypothesis that the ultimate extent of improvements will be best represented by a curve of logistic form. (That the ultimate curve should take on a logistic form, should resemble the letter S when plotted against time or some other measure of effort, is a common belief in the petroleum industry. For us, the 'S' curve of process potential is an hypothesis, to be set against the evidence that we shall collect in the course of our study).

At t(k) the solution to linear programmes (appropriate for t(k)) will indicate the incentives for improvement(s), i.e., the potential benefits. If the sets of constraint equations include measures of the resources consumed in R&D, the solution will also indicate the costs of securing the improvement(s); but this is most unlikely to occur, so *ex ante* estimates of the costs of obtaining improvement(s) will have to be obtained from ancillary data, assuming that even they are available. More likely to be available are *ex post* measures of development costs and (new) equipment prices: those dated shortly after t(k) will give some idea as to the *ex ante* estimate, whereas those dated long after t(k) will give more accurate indications of ultimate (*ex post*) costs and benefits.

Finally, in Figure 2 there are two other extrapolations of the 'S' curve of process potential drawn as dashed, not solid, curves beyond t(k)). If the hypothesis of the 'S' curve is valid, and if the interval t(o) to t(k) is short, relative to the total life of the process, the *ex ante* measure, at t(k), of benefits will give some further indication of the future path, be it the upper (dashed) curve, the middle (solid) curve, or the lower (dashed) curve. The *ex ante* measures thus provide a means of up-dating, in a Baysian manner, the estimate of overall process potential.

Figure 1

Sorts of Comparisons to be Made

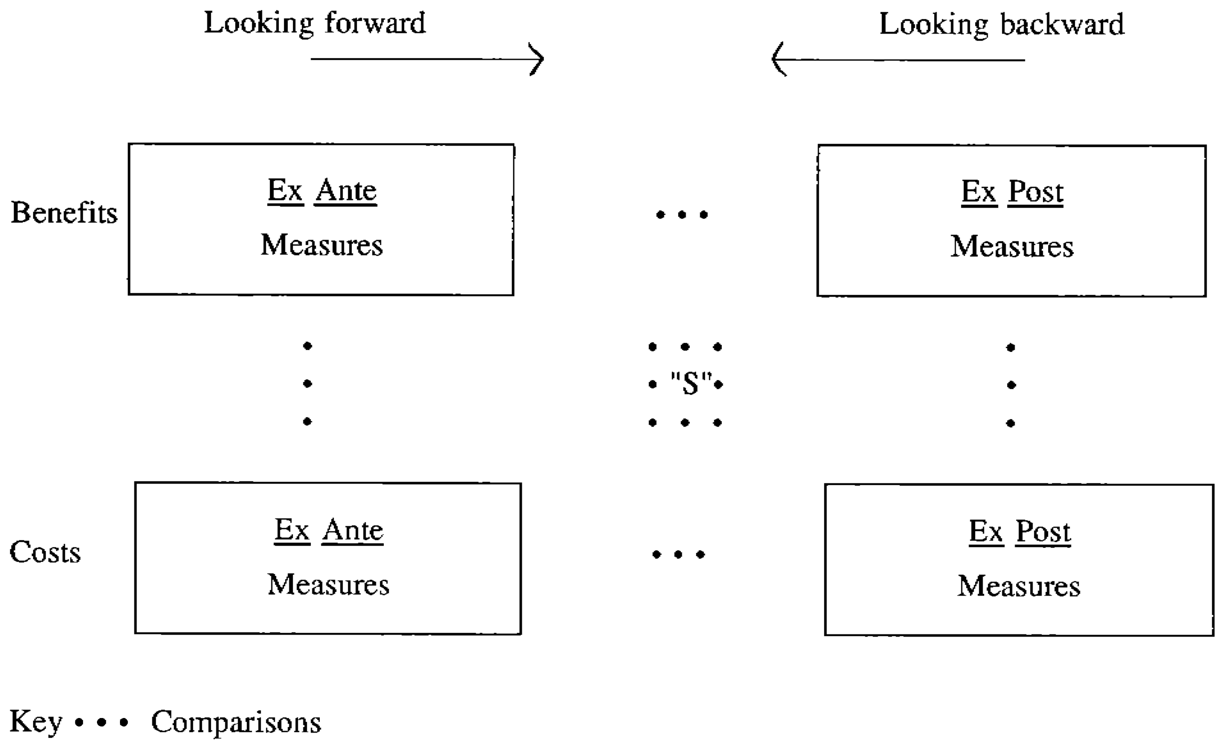
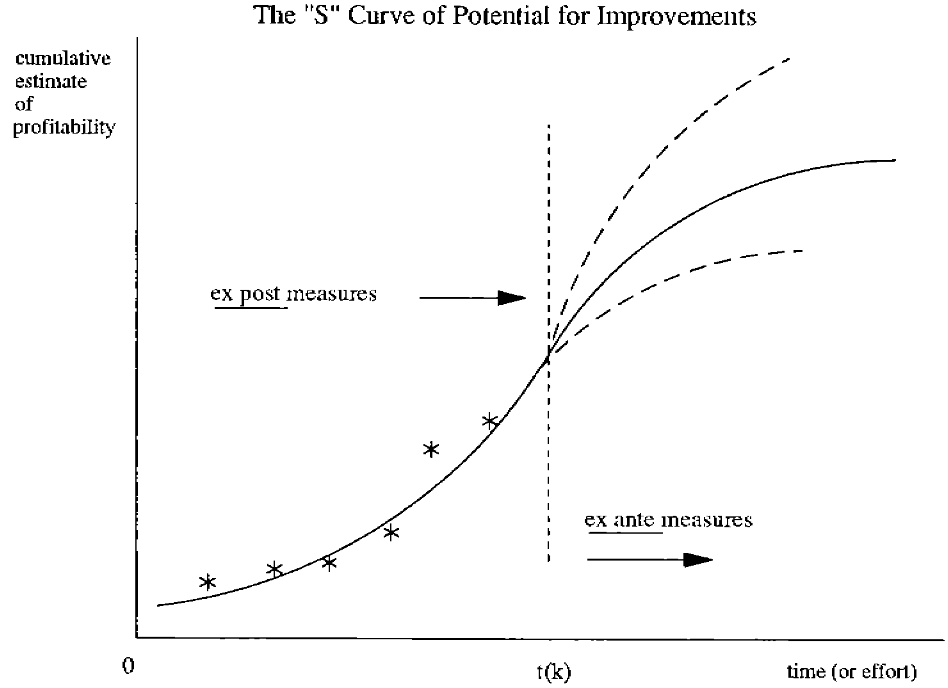


Figure 2

The "S" Curve of Potential for Improvements



BIBLIOGRAPHY

Adams, F.G. and Griffin, J.M. (1972), 'An Economic LP Model of the Petroleum Refining Industry', *Journal of the American Statistical Association,* 67 (September): 542–55.

Charnes, A., Cooper, W.W., and Mellon, B. (1952), 'A Model for Programming and Sensitivity Analysis in an Integrated Oil Company', *Econometrica,* 20 (April): 135–59.

Chenery, H.B. (1949), 'Engineering Production Functions', *Quarterly Journal of Economics,* November, vol. LXIII, no. 4 (November): 507–31.

Chvatal, V. (1983), *Linear Programming*, New York: W.H. Freeman and Co.

Cookenboo, L., Jnr. (1955), *Crude Oil Pipelines and Competition in the Oil Industry,* Cambridge: Harvard University Press.

Freeman, C. (1982), *The Economics of Industrial Innovation,* Cambridge: The MIT Press.

Hadley, G. (1962), *Linear Programming,* Reading, MA: Addison-Wesley.

Hollander, S. (1965), *The Sources of Increased Efficiency: A Study of the Du Pont Rayon Plants,* Cambrige: The MIT Press.

Lester, R.K., and McCabe, M.J. (1993), 'The Effect of Industrial Structure on Learning by Doing in Nuclear Power Plant Operations', *RAND Journal of Economics,* 24/3 (Autumn): 418–38.

Manne, A.S. (1956), *Scheduling of Petroleum Refinery Operations,* Cambridge: Harvard University Press.

Maples, R.E. (1993), *Petroleum Refinery Process Economics,* Tulsa: PennWell.

Nelson, W.L. (1958), *Petroleum Refinery Engineering,* 4th, ed., New York: McGraw-Hill.

Noyce, R. (1977), 'Microelectronics', *Scientific American,* 237/3: 62–9.

Pearl, D.J., and Enos, J.L. (1975), 'Engineering Production Functions and Technological Progress', *Journal of Industrial Economics,* XXIV/1 (September): 55–72.

Rosenberg, N. (1982), *Inside the Black Box: Technology and Economics,* Cambridge: Cambridge University Press.

Symonds, G.H. (1955), *Linear Programming: The Solution of Refinery Problems,* New York: Esso Standard Oil Company.

OXFORD INSTITUTE FOR ENERGY STUDIES
57 WOODSTOCK ROAD, OXFORD OX2 6FA ENGLAND
TELEPHONE (01865) 311377
FAX (01865) 310527
E-mail: publications@oxfordenergy.org
http://www.oxfordenergy.org