DECISION SUPPORT SYSTEMS FOR TARGET SETTING & RESOURCE ALLOCATION IN MULTI-UNIT & MULTI-LEVEL ORGANISATIONS USING DATA ENVELOPMENT ANALYSIS

BY

ANTREAS D. ATHANASSOPOULOS

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF WARWICK

SCHOOL OF INDUSTRIAL AND BUSINESS STUDIES

JANUARY 1995

Αφιερωμένο

σε όλους όσους ξέρουν να δίνουν

ευκαιρίες γιά δημιουργία

TABLE OF CONTENTS

CHAPTER 5 *Centralised target-based planning models in multi-unit & multi-level organisations*

CHAPTER 6 *Decentralised target-based planning in multi-unit & multi-level organisations*

CHAPTER 7 *Diagnostic analysis and decision support for a-priori resource allocation*

CHAPTER 8 *Planning analysis and decision support for a-priori resource allocation*

CHAPTER 9 *Assessing marginal impacts of investments on the performance of organisational units*

CHAPTER 10 *Conclusions and further research avenues*

APPENDICES

APPENDIX OF CHAPTER 4

APPENDIX OF CHAPTER 6

APPENDIX OF CHAPTER 7

APPENDIX OF CHAPTER 8

SYNOPSIS

This thesis is concerned with the development of decision support systems for determining performance targets and allocating resources in multi-unit organisations. These organisations are organised into networks of decision making units (DMUs) that seek to satisfy demand for services in the public sector or attract demand for services in the for profit making sector. Mathematical programming methods in general, and data envelopment analysis in particular are the methods chiefly used throughout the thesis.

The decision support systems sought to address two distinct problems faced by multi-unit organisations.

The **first** is concerned with the allocation of recurrent type of budgets to decision making units which use their resources without interference from their headquarters. This type of problem is called *a-posteriori* decision support and it is addressed by developing a framework of effective target setting. Data envelopment analysis models are developed for setting targets at the DMU and the global organisational levels. Two target-based resource allocation models are then developed seeking to encapsulate alternative organisational structures and objectives of resource allocation, namely *equity, effectiveness and efficiency.*

The **second** case concerns problems where the allocation of resources is made directly to prespecified DMUs. This problem is called *a-priori* decision support which includes a phase of managerial *diagnosis* and *planning*. In the diagnostic phase performance targets for different management tiers are assessed, and systematic procedures of micro level benchmarking are developed. In the planning phase targets for improving the scale size of individual units are assessed, the long & short run viability of the network of outlets is examined and, finally, the marginal impacts of past investments on the performance of DMUs are investigated. The two phases of the decision support system would aid management in making decisions regarding the future of individual DMUs (e.g. investment, expansion, divestment). Application of the method to a network of 154 public houses is incorporated throughout the relevant chapters of the thesis.

Declaration

During the preparation of this thesis in the period October 1989 to December 1994 a number of research and conference papers were prepared as listed below. The remaining parts of the thesis are unpublished.

[1]. "Separating market efficiency from profitability and its implications for planning" (1995) in the *Journal of Operational Research Society* Vol. 46(1) which was also presented at the: First *OR-research* conference in Salford, December 1991. The paper has been written jointly with Emmanuel Thanassoulis and includes some aspects covered in chapter seven of the thesis applied, however, on a different set of data. I hold full responsibility for the ideas put forward and the empirical implementation that followed.

[2]. "Assessing the marginal impacts of investments on the performance of organisational units", forthcoming in the *International Journal of Production Economics*, it has also been presented at the: *Third European conference in productivity analysis,* Louvain, Belgium 1993. The paper has been written jointly with E. Thanassoulis and covers aspects of chapter nine of the thesis. I hold full responsibility for the theoretical model of the paper and its empirical illustration.

[3]. "Performance improvement decision aid systems in retailing organisations using data envelopment analysis", forthcoming the *Journal of Productivity Analysis.* The paper includes some aspects covered in chapter eight of the thesis applied, however, on a different case study.

[4]. "The evolution of non-parametric frontier analysis for assessing performance: Current progress and future directions", *forthcoming in the SPOUDAI an International Journal of Management Science published by the University of Piraeus.* The paper includes aspects covered in chapter two of the thesis.

[5]. "Goal programming and data envelopment analysis (GoDEA) models for target-based multi-level planning: Allocating grants to the Greek local authorities", Warwick Business School Research paper No. 143. The paper includes the theoretical model covered in the chapter five of the thesis.

[6]. "Integrating efficiency assessments with resource allocation in public sector organisations", National Operational Research conference, Birmingham, September 1992.

[7]. "Resource allocation in service organisations with multiple operating units using data envelopment analysis", EURO XII / TIMS XXXI conference, Helsinki, June 1992.

[8]. "Improving resource allocation decisions in the public sector through efficiency assessments", Invited paper in the IFORS 93 XIII, Lisbon, Portugal, July 1993.

[9]. "Performance improvement decision support in organisations with multi-tier management", Invited paper in the best of British section of the ORSA-TIMS conference, Boston, April 1994.

[10]. "Assessing performance and develop cost efficiency & target setting scenarios in "virgin industries" using DEA: The case of Greek local authorities", EURO XIII / OR 36 conference, Glasgow, UK, July 1994.

ACKNOWLEDGEMENTS

There are many people that have influenced the completion of this thesis.

I am wholehartly indebted to my parents (Demitries and Vaciliki) and my sister (Lina) for their financial and moral support during the completion of the thesis. They were there when it was really necessary for me !!

My supervisor Emmanuel Thanassoulis. It is very difficult to describe how influential and beneficial was his presence in preparing this thesis. I recall the tensions and discussions with Emmanuel during the preparation of the thesis with terrifying feelings. I wish, however, more PhD students have the type of supervision I received.

Robert Dyson, for giving advice at the very beginning of the thesis for a secondment in an organisation and for his useful comments regarding the content, style and presentation of the thesis.

Jim Storbeck for his advice, and humorous support while he was visiting Warwick Business School in the period 1991/92. My colleague and friend Joan Ballantine for her patience in reading some of the earlier chapters of the thesis. Chris Strickland for working together at times and days when only the security of the University was around. I would also like to thank research colleagues and friends that gave me good ideas and advice in presenting my research at various national and international conferences: Rachel Allen, Knox Lovell, Larry Seiford, Phil VanEckaut, Finn Forsund, Kostas Triantis.

Finally, financial support from the Science and Engineering Council and Warwick Business School during the first two years of the thesis are acknowledged. I would also want to express my gratitude to the XXX brewery who supplied data and relevant information concerning the empirical part of the dissertation.

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

CRS	Constant Returns to Scale efficiency model (chapter 2)
CTP	Centralised Target-based Planning model (chapter 5)
DEA	Data Envelopment Analysis (chapter 2)
DMU	Decision Making Unit (chapter 1)
DSS	Decision Support System (chapters 5, 6, 7, and 10)
DTP	Decentralised Target-based Planning model (chapter 6)
FDH	Free Disposal Hull (chapter 2)
FPT	Flexible Prioritised Target model (chapter 4)
ISS	Improved Scale Size target model (chapter 8)
MPSS	Most Productive Scale Size model (chapter 2 and 8)
MRT	Marginal Rates of Transformation (chapter 4)
MULO	Multi-Unit Multi-Level Organisations (chapter 8)
MUO	Multi Unit Organisations
PI	Performance Indicators (chapter 1 and chapter 3)
PIMS	Profit Impacts of Market Strategy (chapter 1)
PRISS	Profit Impacts on Scale-size Strategy (chapter 8)
PPBS	Programme Planning Budgeting System (chapter 1)
PPM	Product Portfolio Matrices (chapter 1 and 8)
PT	Prioritised Target model (chapter 4)
ROI	Return on Investment (chapter 1, 7 and 9)
SWOT	Strengths Weaknesses Opportunities and Threat analysis (chapter 1)
VFM	Value for Money Audit (chapter 1 and 3)
VRS	Variable Returns to Scale efficiency (chapter 2)
ZBB	Zero Base Budgeting (chapter 1)

PROLEGOMENA & THESIS STRUCTURE

Performance measurement, target setting and resource allocation are key issues in the managerial agenda of for-profit and not-for-profit organisations. It is widely recognised that organisations' viability rests more than ever on their ability to achieve and sustain high levels of internal and external performance at all levels of their operations.

Performance measurement, however, is not a set of numerical figures obtained mechanisticly by a particular function of the organisation or by some external consultants. In this thesis, performance is viewed as an integrated part of the organisational operations. The latter implies coordination and integration between performance measurement and decision making (i.e. resource allocation).

The integration of performance measurement and resource allocation in multi-unit organisations is a key theme of the thesis. This is pursued via the development of decision support mechanisms that seek to incorporate managerial preferences, from different levels of the organisational hierarchy, in developing systematic processes for setting targets and allocating resources.

The thesis is organised in three thematic sections of ten chapters in total. **Section one** (chapters 1 and 2) defines the problem area of the thesis, whilst a literature review reports on methods for resource allocation and performance measurement. **Section two** (chapters 3, 4, 5 and 6) is concerned with the development of decision support systems for target setting and resource allocation in *a-posteriori* decision making. That is, DMUs of the organisation are responsible for deploying resources allocated centrally by top management in the form of budgets. **Section three** (chapters 7, 8 and 9) contains decision support systems for *a-priori* decision making. That is, DMUs are allocated resources from their headquarters to undertake specific tasks (projects). The thesis concludes in chapter 10 with an overall assessment of its achievements and a discussion of possible avenues for further research. The structure of the thesis is discussed in more detail next.

SECTION 1
MANAGING MULTI-UNIT ORGANISATIONS: CURRENT PRACTICE & FUTURE DEVELOPMENTS

Chapters one and two develop the research agenda of the thesis and also review the state of the art methodologies in the field of our research. As research objective we have envisaged the development of decision support mechanisms that will enable management to link *control* and *planning* mechanisms in multi-unit organisations. C*ontrol* and *planning* will be operationalised using target setting and resource allocation respectively.

Chapter one outlines the research questions addressed in the thesis. Illustrative examples of multi-unit organisations are used as vehicles for introducing the concepts of control and planning. The general definitions concerning performance measurement and resource allocation are followed by a review of the state of the art methodologies for resource management with particular emphasis on their ability to capture performance characteristics. The chapter leads into developing a broad classification of multi-unit organisations and also the nature of their *control/planning* problems which will be explored further in the subsequent chapters.

Chapter two focuses on the state of the art methods for assessing performance and it is organised in two parts. In the first part the political, accounting, economic and management science dimensions of performance measurement are discussed. Frontier analysis is then presented as an attempt to synthesise these alternative views. The chapter proceeds analysing the technical framework of frontier analysis with particular emphasis given to issues that will be used in the main part of the thesis.

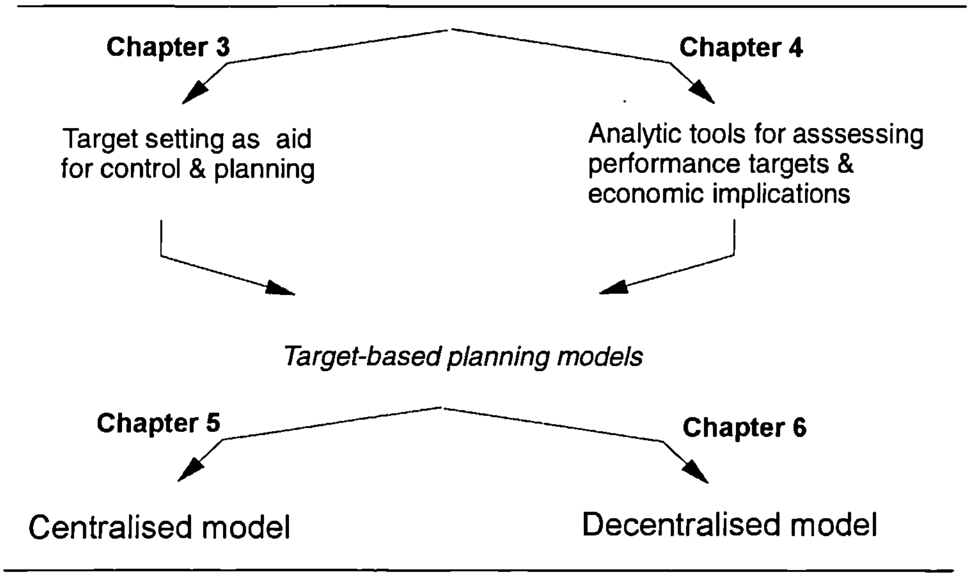
SECTION 2

TARGET SETTING & RESOURCE PLANNING IN A-POSTERIORI RESOURCE MANAGEMENT

In the research agenda discussed in chapter one a distinction is made between "market" and "relative need" oriented cases of performance measurement and resource planning. This section focuses on the development of performance based decision support methods for "relative need" cases. In need based resource allocation the management of individual operating DMUs decide on the use of centrally allocated global resources. This relative flexibility of DMUs, to allocate resources to various activities according to their discretion, creates an ***a-posteriori decision making*** framework. The development of decision support mechanisms for this type of problem constitutes the research agenda of this section.

Section 2 is divided into four chapters as can be seen in the pictorial representation below.

A-posteriori decision making framework



In **chapter three** it is argued that in a-posteriori decision making performance measurement and planning mechanisms should be linked. This is not feasible, however, as performance measurement and resource management are not compatible by being parts of distinct organisational processes, namely *control and planning*. Target setting is introduced as an effective instrument for linking control and planning processes in MUOs. The enhanced role of target setting needs to be supported by a set of target setting principles. More advanced operational models need to be developed in order to address the principles of target setting.

Chapter four draws upon the conclusions of chapter three and develops frontier based models to assess performance targets in line with the principles of target setting. A prioritised target framework model is put forward for assessing performance targets at the individual activity centre and also at the global organisational level. The managerial and economic implications of these models are explored further using a numerical example.

Chapter five seeks to pursue the resource management objective of the decision support system. A target-based planning model is put forward anticipating the links between target setting and resource allocation in multi-unit multi-level organisations. This was done having found limited support from the current literature on multi-level planning to address this type of problems. The planning model in this chapter is appropriate for organisations that adopt a centralised behavioural system of decision making. Issues related with the characteristics of this model and its ability to satisfy the principles of target setting as well as the planning objectives of MULOs are also discussed.

Chapter six concludes the a-posteriori decision support section by putting forward an alternative framework for target-based planning. This framework seeks to capitalise on the case of the decentralised approach to decision making. Crucial issues here are the representation of individual activity centres on the planning process, and also the coordination of the different levels of decision making in the planning process.

SECTION 3

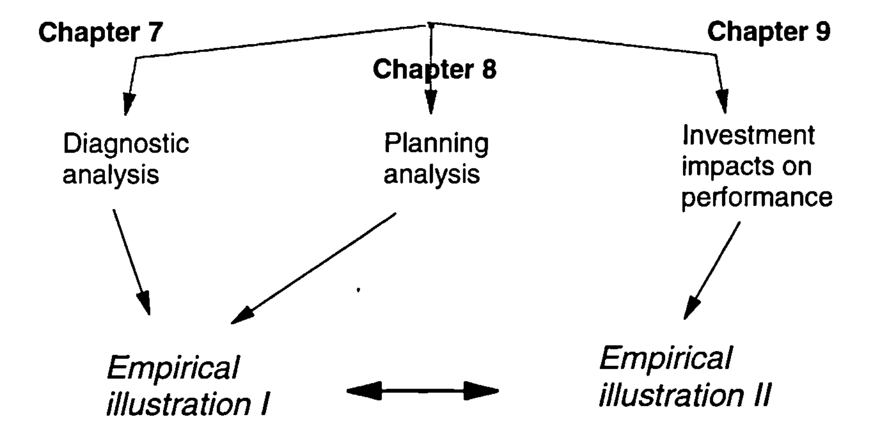
TARGET SETTING & RESOURCE PLANNING IN AN A-PRIORI RESOURCE MANAGEMENT

In the previous section (chapters three, four, five and six) a decision support framework was developed for assisting managerial control and planning in not-for-profit multi-unit organisations. This type of problem was called *a-posteriori* decision support as individual operating DMUs were allocated resources without project specific mission, (e.g. recurrent grant to a university).

This section (chapters seven, eight and nine) seeks to investigate issues of control and planning in cases where DMUs are selected *a-priori* for allocating resources. This is a problem often met when capital investment decisions are made in for-profit or not-for-profit organisations. The nature of capital investment requires development of *a-priori* decision support framework where investment opportunities are identified, evaluated and thus capital is allocated. The term *a-priori* indicates the *project-based* character of capital investment decisions. It is argued that the development of effective *a-priori* decision making processes should provide *diagnostic* and *planning* related management support.

A pictorial representation of how this section is organised is given below:

A-Priori decision making framework



Chapter seven is concerned with ***diagnostic*** problems of performance in for-profit making multi-unit organisations. A key issue of the diagnostic analysis, is the definition of market and cost efficiency as separate, although related components of performance. Market efficiency is used as a tool for assessing the extent to which for-profit making DMUs utilise their market potential and generate sales. The presence of a two tier management in multi-unit organisations makes it necessary to distinguish the assessed market efficiency for each tier of management. This chapter is also concerned with the identification of "role operating practices" that can be used as benchmarks within organisations. These ideas are illustrated using a set of 154 pubs from one of the major UK breweries.

Chapter eight seeks to develop decision support mechanisms with planning orientation. This is pursued by developing models for assessing improved scale size targets which is the key issue of the chapter. It concerns the elimination of DMUs' inefficiencies due to the wrong scale size of operation. The long run viability of outlets is assessed using their profitability and market efficiency scores. Finally, the effectiveness of service mix provided by individual outlets is another important element that would affect planning decisions in organisations.

In chapter nine we seek to advance the current methodology for assessing the effectiveness of capital investment decisions on the infrastructure of individual DMUs. As the use of accounting measures is by no means sufficient for assessing the impacts of infrastructure investments, a DEA based methodology is developed to address the question more effectively. The estimated **investment effectiveness** indices seek to evaluate the extent to which past investment policies have succeeded in improving the market efficiency of individual DMUs. The latter in conjunction with the planning market efficiency targets estimated in chapter eight can provide invaluable information on the selection of the most prominent DMUs during the capital investment process. The method developed was applied on the set of pubs used in the previous chapters.

Chapter ten is the concluding chapter of the thesis. An overall evaluation will be made of the extent to which the thesis objectives as stated in chapter one were achieved. A validation framework will be proposed for assessing the usefulness of control and planning models developed in the thesis. Finally, the chapter proposes avenues of further research.

Chapter 1

Developing a research agenda of management control and planning in multi-unit organisations

1. Introduction

Multi-unit organisations (MUOs) are perceived within general organisational clusters regarding the nature of their services e.g. food sector (grocery chains), financial services (bank branches), public service sector (schools and hospitals), utilities (electricity generating plants), etc. The practices followed for assessing performance and managing resources are very often determined by the characteristics of the industrial sector an organisation belongs to. In the Brewing industry, for example, most companies use periodic profitability figures (e.g. profit per ft^2 of bar area) in order to monitor outlet performance while in the telecommunications industry the cost per connection has widespread applicability as a key measure of performance (see Manzini and Thalmann (1994)).

Despite the presence of alternative methods for assessing performance there is always a core set of indicators which are considered by organisations as the most important. As the role of performance measurement is rising significantly over time, some questions can be posed regarding the appropriateness of the traditional performance measurement methods to cater for the demanding features of the changing economic/business environment. The linkage between performance measurement and decision making mechanisms (e.g. resource allocation) is one of the issues that has gained considerable popularity during the last two decades[1]. This issue is one of the primary motives of this research, whose principle observations can be stated as follows.

- *The development and implementation of synergistic processes between performance measurement and resource planning can aid the effectiveness of control and planning mechanisms in MUOs.*
- *Methods currently used for assessing performance and allocating resources in MUOs are not always compatible. The development of synergies between the two would require methodological enhancements.*

The thesis ultimately develops and implements models and processes which link performance measurement and resource management in MUOs.

The remainder of this chapter is organised as follows. An introductory analysis of the nature and basic features of MUOs is followed by a discussion of the concepts of managerial control and planning applied in MUOs. In the context of this study performance measurement is considered as the basic component of managerial control, whilst resource management is considered as the basic function of managerial planning. Most resource allocation methods developed since the evolution of scientific management by Taylor (1911) sought to improve organisational performance. The extent to which performance is incorporated effectively within these resource management methods is examined using evidence drawn from the literature. A research framework focusing on resource management and performance measurement in MUOs is then developed revealing topics to be considered in the remainder of the thesis.

2. The essence of multi-unit organisations (MUO)

During the last 20 years there has been a considerable shift in Western Economies and in the UK in particular towards service sector economic activities. New patterns of interactions between customers and suppliers have emerged following the growth in service sector activities. Organisations sought to be as close as possible to their customers and customers' needs. Business strategies (e.g. gigantism in the size of International Business Machines) that led firms successful in the past are theoretically and practically inappropriate at present. New types of organisational structures (e.g. reduction or elimination of middle management

---

[1] In the UK, for instance, the Conservative governments (1981 -) seek to link pay rises in the public sector with direct productivity improvements.

in large corporations) emerge as we move towards the end of the century (see Tom Peters: Liberation management, 1992).

An interesting type of organisational structure is that of the multi-unit organisations which share, as a common feature, the presence and management of multiple activity centres. In the profit sector one can find retail stores organised into networks, manufacturing plants with multiple sites, etc. Activity centres in the private sector tend to be called "profit units" whilst in the public sector "cost or service units". However, this is not always the case as for instance bank branches are considered more as activity and less as profit centres. Illustrative examples of MUOs can be found in different facets of economic activity: Commercial banks, Public houses and Restaurants, Supermarket chains, Insurance companies, Sales forces, Air transport companies, Hospitals, Educational institutions, Patrol stations, Post offices, Police departments, Manufacturing plants, and Electricity and water production/distribution units.

For-profit and not-for-profit MUOs, have followed a steady pattern of expansion (and often over expansion). In the not-for-profit sector this was a result of the interpretation given by successive central governments to the tax payers' pressure for higher quantity and quality of services (equitable provision of services) in the post second world war era. It was in the early eighties that the over expansion trend was challenged for its inefficiency but more importantly its macroeconomic implications. In the for-profit sector on the other hand the annual reports of many corporations give particular emphasis to the increasing number of their outlets as a sign of strength. Excess capacity in many for-profit making industries (e.g. bank branches, patrol stations and supermarkets) has resulted in marginal revenues that do not exceed the marginal costs.

MUOs have many similarities in their structural and operating characteristics:

- Producing/delivering multiple goods and services,
- Are organised into spatial networks that are linked hierarchically,
- Individual units of MUOs operate under internal and external "market"[2] conditions.

---

[2] The case of developing internal markets has recently become the flagship for UK Governmental reforms in the public sector.

Different levels of management are involved in the administration of individual decision making units (DMUs) depending on the type of decisions concerned. For example, a bank branch is affected by its own manager, by the regional planning management and by the central headquarters of the bank for policy, marketing, resource levels and other issues. The presence of intermediate levels of decision making can lead to tradeoffs and internal conflicts between the different levels of management. At the most extreme, one can argue that in certain cases (e.g. capital investment decisions) MUOs operate as internal markets where the different levels of administration "compete" for higher authority and access to decision making. The immediate implication of this is that the assessment of a DMU's performance should accommodate the different levels of management that have a direct or indirect impact on its operations.

2.1. For-profit and not-for-profit MUOs

Undoubtedly, for-profit organisations are affected by the competitive market in which they operate. The fundamental objective of a profit making enterprise is to penetrate its market and generate profit. Profit, however, is not the only criterion that dominates resource allocation decisions. Galbraith (1977), among other influential writers, notes the existence of a wide spectrum of criteria regarding the objectives of profit making organisations. For example, shareholders' satisfaction, market share, long range stability, return on investment, and adaptation to technological innovations and economic shake-ups.

Not-for-profit organisations are mainly involved in the delivery of public and social goods (health, education, social security, emergency services and charity services). Resource allocation, decision making and performance measurement in these organisations is always a composite function of governmental policies, citizens' expectations and management's reactions.

The multitude of organisational objectives does not allow for "*black and white*" distinctions between for-profit and not-for-profit organisations. Generally, it can be argued that MUOs, while maintaining autonomy (due to their different nature) in setting their goals and objectives share similar resource management problems. For instance, a for-profit making establishment may allocate its resources towards increasing/sustaining the demand for its services/products and generate profits. On the other hand, a not-for-profit organisation uses its resources to meet the demand for services (quantity/quality) at the minimum possible cost. In both cases there is an element of choice from central management in order to get the highest possible returns from its decisions to allocate resources in one or the other way.

Management is always expected to make decisions towards maximising the organisational "welfare". Welfare is conceptualised as the satisfaction of organisational aims (mission). These aims are expressed through the development of objectives; objectives are then transformed into goals; goals lead to organisational targets and overall they constitute a welfare structure (function) of an organisation. The maximisation of this welfare will be supported in the thesis by the development of effective resource management and performance measurement methods which are parts of the more general concepts of planning and control in MUOs.

2.2. Managerial control and planning in MUOs

Managerial control and planning are inextricably linked concepts in the operation of organisations. There is a host of organisational behaviour, business policy and accounting literature concerned with issues of organisational control and planning.

There is no unique definition describing *managerial control,* Hogler and Hunt (1993). However, most research seems to agree that Antony's (1965) definition "control is the process of assuring that the organisation does what management wants done" is universally accepted as a starting point. Theories of control have been developed along the lines of organisational and agency theory approaches. The organisational approach, Ouchi (1979), recognises a performance measurement and a social based control strategy.

In this research the performance led strategy has been adopted and therefore control encompasses **output** and **behavioural** issues. The representation of behavioural issues in assessing performance is discussed, in detail, in chapter four where the extent to which goals and targets are achieved is assessed incorporating managerial priorities.

The agency theory approach draws upon the economic and accounting theory of control. The fundamental concept is that of contracting, which is based on the dual scheme between *principals* and *agents.* Following the principal-agent paradigms control is realised via behavioural and outcome based strategies.

Comparison of agency theory with the organisational approach reveals similarities and differences. As Eisenhardt (1985) argues, both agency and organisational approaches are rational, efficiency based, and both distinguish between behavioural and outcome based control. Insofar as the differences between the two theories are concerned, Eisenhardt (1985) notes differences in costs, rewards, information, and uncertainty.

In summary, the organisational and agency approaches are complementary. The organisational approach emphasises:

- The importance of task characteristics to the choice of control methods,
- The existence of social control as an alternative to performance based control.

Agency theory adds to the organisational approach emphasis on:

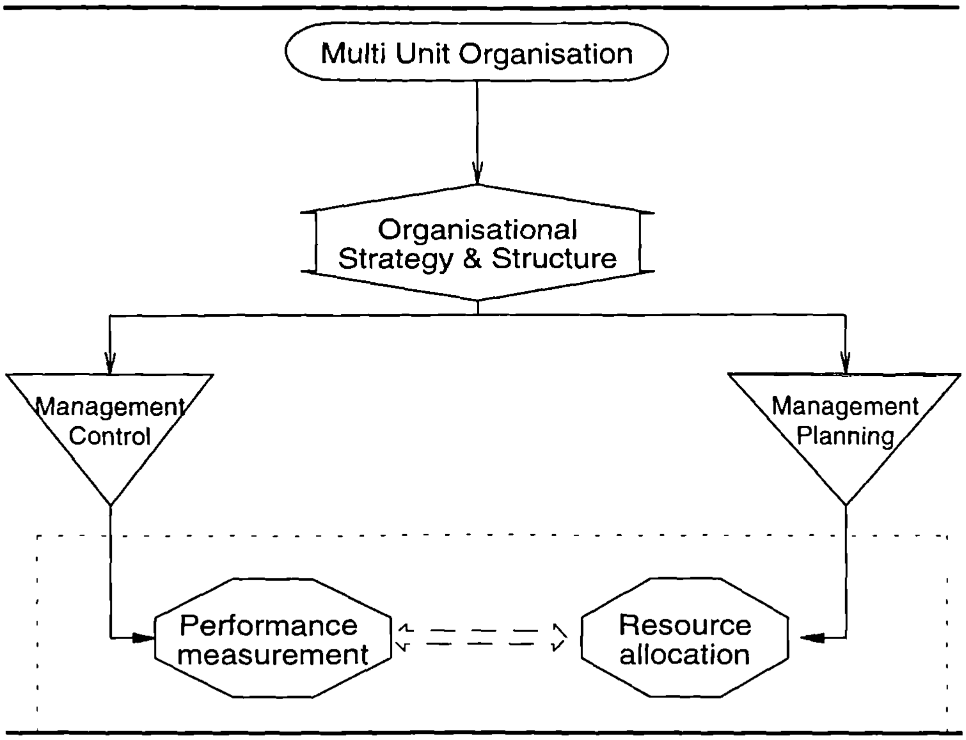
- Information systems
- Uncertainty
- Costs
- Rewards

Management control in isolation of the methodology adopted for its support is an important aspect of organisational design, Eisenhardt (1985). Rotch (1993) defines a conceptual framework of management control using a set of five components: Performance, Strategy, Organisational structure, Direction and Motivation. Based on this framework Rotch argues the importance of considering the interrelationships between these components in a management control system.

A unified framework of managerial control and planning can be found in Figure 1.1. This framework emphasises the role of performance measurement as a managerial control component and also introduces management planning as an internal component of the control process. The joint consideration of management control and planning systems will be the basis for investigating the research hypotheses of this thesis.

Figure 1.1

Managerial control & planning in MUOs



There is a number of assumptions portrayed in Figure 1.1 concerning management control and planning. Organisational strategy and structure are considered as the starting points of MUOs management which is then employs control and planning mechanisms for realising the strategies. Performance measurement and resource allocation are among the most typical perceptions of management control and planning mechanisms in MUOs. Furthermore, in Figure 1.1 an extended conception of management control is advocated, which enhances the role of control from a simple feedback mechanism towards an integrated process that interacts with the very nature of organisational design. This type of control systems seek to relax the traditional view that management control follows on and is inferior to the organisational design process.

A dialectic consideration of organisational control and planning processes would need to expand the traditional performance measurement and performance reporting features of control. It is argued that this enhancement of control mechanisms can be achieved by linking performance measurement with target setting and resource allocation processes.

Performance measurement is a **multi-disciplinary** concept which allows different disciplines e.g. *political, organisational behavioural, economic and management science* to adopt their own perspectives, with profound implications, in defining and assessing performance. These alternative perspectives can also be found when analysing the development of planning (resource allocation) processes in MUOs. The effects of alternative perspectives of performance assessment are discussed in more detail in chapter two.

3. Managing resources in organisations: Concepts & current practice

The operation of organisations is based on the use and deployment of resources. There is a variety of tangible (e.g. human resources) and intangible (e.g. information) factors that can be considered as resources. It is argued, however, that resources (inputs) have **no value** unless organised into **systems** which ensure that resources will be transformed into goods and services (outputs). The design of effective systems for allocating, managing and controlling resources determines the organisational success. Issues concerned with the allocation and mix of resources and also with the relation between resource management and organisational structure are discussed next.

Decisions concerning resource allocation deal with two principal issues:

i. The broad issue of **how** resources should be **allocated** between the various functions, departments, divisions, and separate units.

ii. The more detailed considerations of **how** resources should be **mixed** and ultimately deployed within organisational units.

In the parlance of resource auditing resource mix reflects the "technology" employed by individual units and/or organisations. As Gold (1981) emphasised, an on-going scientific debate exists concerning the significance of input/output mix and scale/size problems in the neo-classical production theory. Chapter 2 will show that the economic approach for assessing efficiency assumes that the input/output mix of the assessed units/firms is to be preserved.

The amount of allocated resources among different operating functions is guided by the achievement of strategic/corporate organisational plans. The distinction between the amount of used inputs ***versus*** the optimal mix of inputs reflects the fundamental economic definitions of performance (technical ***and*** allocative efficiency, Fare (1988)) which are analysed in chapter two.

The significance of resource management can be seen in conjunction with many other managerial issues, for example target setting, corporate and strategic planning, Ansoff

(1987) and Argenti (1989). It can be argued that resource management is concerned with the capacity constraints whilst corporate and strategic planning reflects policy constraints in view of the global organisational profile. This is portrayed graphically in Figure 1.2.

Figure 1.2
Policy & capacity domains in MUOs

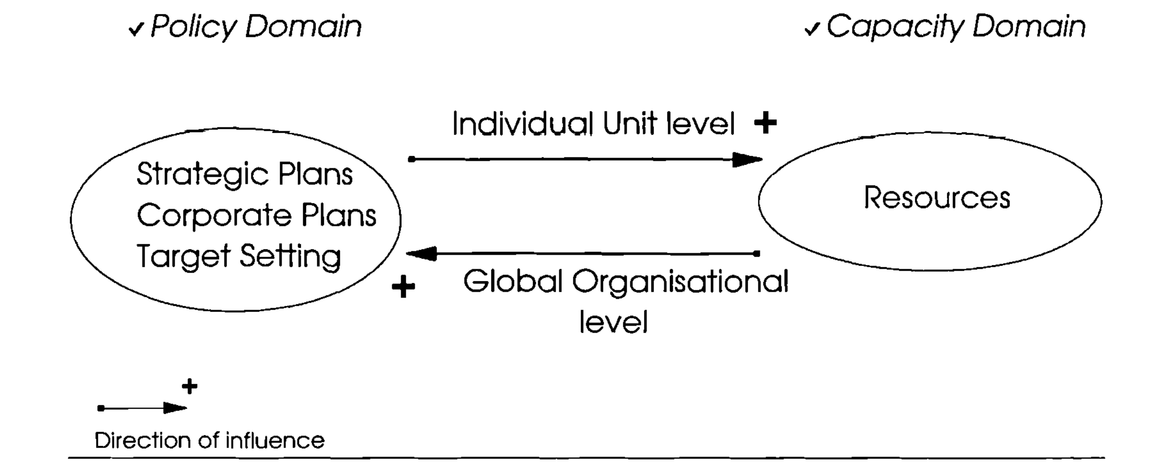


Figure 1.2 portrays the mechanism guiding the association between policy and capacity constraints in a MUO. Resource availability[3] influences the decision making process towards the **global** organisational achievements. In contrast, the policy constraints affect the allocation of resources at the **individual** unit level. The two issues, *deciding* at the *macro* and *implementing* at the *micro* level need to be addressed separately. This is due to the fact that the objectives/criteria for making decisions at the micro and macro level are different. Macro level resource commitments encapsulate a "feasible" version of the strategic vision of the organisation due to capacity limitations. On the other hand, micro level resource commitments facilitate the achievement of strategic and corporate targets with reference at the individual operating unit. The case of global and unit level decision making will be considered in more detail in chapters five and six with operational models that will address both problems simultaneously.

To facilitate the disentanglement of **capacity** and **policy domains** let us consider a commercial bank. The total amount of capital to be allocated amongst bank branches in a time period forms the capacity domain of the bank. In other words, the bank's central

[3] Even in the case of loans and external financial support, at the stage of planning an organisation has a given amount of resources to allocate.

management has a given amount of capital to allocate among a set of operating activity centres, namely bank branches[4]. The achievement of corporate objectives of the bank are customised per individual branch and supported by appropriate resources. Illustrative cases of ***short run*** objectives concern profitability achievements; capital investment; the level of controllable expenses, etc. The ***long run*** objectives of the bank are concerned with its market share; the type of targeted clients; the number of operating branches, the policies regarding its major competitors, etc.

The capital available for investment will, effectively, influence the bank's expectations concerning the achievement of its ***short*** and/or ***long*** run objectives. For instance, the extent of redecoration in the branches' network is a function of capital availability. However, the actual implementation of the allocation process follows the opposite direction. The aim of accomplishing the policy objectives of the bank will drive the allocation of resources at the individual bank branch's level.

The previous example highlights the two way communication regarding the association between target setting and resource allocation in MUOs. Resource management overshadows target setting **at the global organisational level**. On the other hand in the **management of individual operating units** the allocation of resources is guided by the targets' achievement. The two way association between target setting and resource allocation creates trade-offs within multi-unit and multi-level organisations that will be discussed in more detail in chapters five and six.

3.1. Objectives for unit level resource allocation

The allocation of resources between operating entities is a major concern of corporate and strategic planning, Johnson and Scholes (1988). Resource allocation decisions have to be compatible with the <u>short</u>, as well as <u>long</u> run organisational welfare. Organisational objectives need, therefore, to guide the resource allocation processes. **The term objective indicates a preferred direction of movement for an organisation**[5], Keeney and Raiffa (1976). To measure and stimulate the achievement of individual objectives organisations

---

[4] Generally speaking organisations decide at the beginning of each year on a level of expenditure. It is then a distinct issue to allocate resources to various organisational functions.

[5] A very interesting debate can be found in the multiobjective programming literature regarding on the nature of the organisational objectives in Operational Research, notably optimisation versus satisfaction, Achoff (1982) and Zeleny (1982).

use sets of *criteria*. For example, the objective of profit maximisation can be reflected using as criteria return on investment, return on equity, return on assets, residual income, etc. An objective for maximising the quality in the provision of health services can be reflected using patients' perspectives, recovery rate and quality adjusted life years (QALY).

A set of aims affecting resource management at the individual unit level is as follows:

- Improve the short-run achievements
- Reward the good performers over time
- Support underperforming units with high potential

Discussion for each attribute is provided below.

3.1.1. *Improve the short-run achievements*

The time horizon of achievements sought by organisations is affected by the nature of their stakeholders. A public limited company (plc), for example, may seek short run achievements in order to increase its share price in the stock market. Coates (1989) argues that in the post-1980 period in the UK the short run organisational performance has proven to be a highly significant factor guiding decision making. Short-termism in isolation can be a very dangerous practice with very adverse implications on the long run viability of organisations.

3.1.2. *Reward the good performers over time*

This attribute relates both to the short and the long run success of an organisation. Organisations seek to identify units with good operating practices and use them as benchmarks. Benchmarking has become a very fashionable tool of business policy, (see Journal of Business Strategy, (1993)). However, the identification of benchmark units is not sufficient if it is not supported by corresponding motivation mechanisms with the following characteristics.

a. Quantity

A good performer is a potential candidate (not always[6]) for being given additional resources.

---

[6] As Athanassopoulos and Thanassoulis (1991) mentioned, there are cases with units operating efficiently but having reached a saturated level in their performance due to exogenous factors.

b. Effort

Rewarding the good performers provides stimuli for the underperforming operating units to improve their efficiency.

Evidently, "rewarding the good performers" has a "means to an end" character concerning the accomplishment of organisational objectives. In effect, rewarding "benchmarks" is a **necessary** but not **sufficient** condition for improved performance. Penalising the "bad performers" on the other hand, which could be seen as a natural by-product of the rewarding mechanism, could worsen the position of poor performers.

Needless to say the definition and identification of *"good and bad performers"* is a very important issue. MUOs are currently employing traditional profitability ratios in the private sector or performance indicators in the public sector for identifying good and bad operating practices. Chapters seven and eight discuss in more depth the limitation of these methods to assess performance and identify benchmarks.

3.1.3. Supporting underperforming entities with high potential

As mentioned earlier on, the management of MUOs allocates resources in order to achieve organisational objectives. A for-profit organisation, for example, allocates resources to increase demand for marketed products/services. To pursue this task management is called upon to identify individual units with products and services with the highest marketable prospects and then allocate resources. On the other hand, a "regulated" agency (public sector) allocates resources in order to satisfy demand for services under the regulation restrictions imposed by government or other regulating bodies (e.g. equitable provision of services, allowable price increases, allowable rate of return). Allocated resources are used to ensure sufficiency, efficiency and high quality in the provided services in a sector, where the demand is exogenously determined and there is legislation for its satisfaction (e.g. policing, education, health, etc.).

Undoubtedly, in both for-profit and not-for-profit organisations the available resources are scarce. This creates a problem of **choice** in the actual allocation of resources. *Should more resources be allocated to efficient units or should they be used to support underperforming units to revive their performance ?* This dilemma stems from the fact that it is very difficult to forecast the outcomes from the allocation of extra resources to an operating unit. For example, the allocation of extra resources to efficient units can increase production levels due to their very good operating practices. Equally likely, however, extra resources allocated to efficient units can be underutilised as the unit's performance can be "saturated" due to exogenous factors (e.g. value added achievement at schools may prove to be bounded by the very poor socio-economic status of the pupils' families). Similar arguments hold in the allocation of resources to inefficient units. **Therefore, the current level of performance, in itself, does not indicate whether extra resources need to be allocated to a unit.**

It is argued that the allocation of resources should be given priority towards the units with the highest "potential" for improving the overall performance of the organisation. The term "potential" has a twofold interpretation. Firstly, it measures the fertility of the environment in which a unit operates in (e.g. market's size for a profit making unit) and secondly, it reflects the internal ability of the unit to operate effectively (e.g. operating practice). Assessment of the external potential of a unit is particularly useful in the allocation of resources to profit making units and it will be explored further in chapter seven and eight. On the other hand, the tradeoffs between the allocation of resources to individual units and the improvement of the global performance of the organisation are of particular importance to non-profit organisations (see chapters five and six).

In summary, rewarding the good performers as well as supporting the underperforming units with high potential creates managerial trade-offs. Moreover, the achievement of optimum financial or fiscal type (e.g. market share) figures, in the short run could be conflicting with the long run achievements[7] of organisations. Finally, resource allocation needs to consider the potential incompatibilities and trade-offs between the allocation of resources in view of individual units' needs as opposed to those that seek to maximise the *global* organisational welfare. This issue is pursued in more detail in chapters five and six of the thesis.

3.2. Resource management methods & performance

Management accounting literature, Ashton *et al.* (1991) argues that resource management ought to be seen in line with organisational control and planning and not just as "routine" based periodic job.

---

[7] As mentioned, it is very common in the post-1980 period to see organisations sacrificing the long run success, in favour of the short run financial targets.

Undoubtedly, resource allocation is the "bread and butter" for different functions of an organisation. Budgetary planning; investment appraisal; portfolio investment; maintenance expenses and recurrent grants are illustrative cases of resource management in MUOs. This is an area of intense research by the accounting, economics, and management science disciplines during the last 30 years. A plethora of general processes have been developed and applied in various real life organisations e.g. Incremental Resource Allocation, planning Program Budgeting System (PPBS), Zero Base Budgeting (ZBB), Quasi markets (contracts) in the public sector, Profits Impacts on Market Strategy (PIMS), Product Portfolio Management, (PPM) evaluating the resource allocation process in the public and/or private sector of the economy. Despite the presence of numerous models for resource allocation found in the literature ***the challenge of performance measurement and its enhancement as an integral part of resource allocation is yet to be addressed sufficiently.***

A synopsis of the state of the art methodologies of resource management follows next. It draws, mainly on the limitations of traditional resource management methods to encapsulate performance measurement characteristics. This analysis **highlights the necessity for enhancing the current perceptions and practices of resource management to integrate it with performance measurement.**

Undoubtedly, **incremental** resource allocation is a common practice for a large number of profit and non-profit organisations. The term "incrementalism" refers to a pattern of marginal change in final allocation of capital relative to some base, which is frequently the previous year's capital allocation, Davis and Dempster (1966). The simple to use and the pragmatic character of incrementalism are necessary but not sufficient conditions for effective resource management. The main drawback of this methodology emanates from the fundamental assumption that the past decisions for allocated resources were used efficiently without any allowances for inefficiency and/or mismanagement.

The dissatisfaction with the incremental method led to the development of alternative methodologies for managing resources like the **Program-** planning-**Budgeting System** (PPBS)[8]. The key feature regarding PPBS is the "program" element as being the subject of

---

[8] PPBS have originated (1969) from Texas Instruments Inc. and transferred by Mac Namara, at a later stage, in the public sector

the resource allocation appraisal in an organisation. For instance in the context of, say, Local Authorities the tax collection service requires collaboration of staff operating in different departments. Under PPBS rules the "program" will be evaluated as if it was a coherent "program" providing the given service. The very demanding procedures (timewise and managerial effort) for implementing PPBS is one of the main drawbacks of the method in real life problems, Coates *et al.* (1989).

In an attempt to overcome the difficulties of the PPBS a new system, namely Zero-Base Budgeting, has been introduced. In essence, **Zero Base Budget** (ZBB) is a decentralised mechanism by which management can closely examine and reconsider the resource levels used by operating units. Departmental activities are organised into decision packages which are then prescribed functions, goals, and costs. Each decision is evaluated and ranked in the light of cost-benefit analysis and it is repeated from a *zero base*[9] every year. The method succeeds in focusing managerial attention on reviewing the organisational functions and their costs and benefits. In practice, the management responsible for implementing ZBB has proven very indecisive in articulating preferences over alternative decision units which made the method very time consuming. Finally, the strong judgmental/qualitative character of the method does not allow for the measurement of any direct impacts on the productivity of the operating units, (Sherman (1986)).

A relatively new methodology of resource management in the public sector in the UK is the contract or **quasi internal markets** with main area of application (at present) the National Health Service in the UK. Resources are allocated by district health authorities to hospitals with the most "attractive" bids concerning particular health services. The basic assumption is that the quasi-competitive market will force hospitals to increase efficiency and therefore reduce the costs of running the National Health System (NHS). Despite the embryonic stage of the system it has already surged a very serious scientific and political debate regarding the evaluation and appropriateness of the quasi-market experiments, Ferlie (1993). The early signs from the application of the quasi-market system indicate increases in efficiency whilst strong concerns have been raised regarding the deterioration in equity and possible effectiveness of the provided services. The importance of these objectives of resource management are discussed in more detail in chapter four and six of the thesis.

---

[9] The non-historical considerations of the ZBB constitutes the main difference from the PPBS.

The **Profit Impacts on Market's Strategy** (PIMS) system is a popular tool for resource management and strategy evaluation of profit making organisations, Schoeffler (1974). PIMS seeks to combine information from a wide sample of organisations with ultimate scope to provide decision support to companies included in the sample of the system. Nowadays, a whole variety of databases with purposes similar to PIMS can be found around the world. However, the use of very simplistic analytical tools for obtaining statistical forecasts has been criticised as the main limitation of these systems. As it will be argued in chapters seven, eight and nine, the limitation of these models emanates from their inability to disentangle different components of organisational performance (e.g. distinguishing responsibility of different level management) and also from the lack of differentiation between *diagnostic* and *planning* focused analyses (see Epstein and Henderson (1989)).

Strategic resource *planning* has been given great academic and consulting research attention in the late 1960s. Illustrative examples are the Boston Consulting Group (BCG) model, Henderson (1973), the Industry Attractiveness and Business Strength model, Hax and Majluf (1983), the Marakon's model (1980). These models view firms as a portfolio of businesses, each one offering a unique contribution to growth and profitability. The methods adopted are based on the development of product portfolio matrices classifying individual businesses into business clusters with different planning implications. The growth-share matrix systems have had a major contribution to strategic resource planning providing a basis for systematic analysis. Major achievement of this debate is the realisation of the multidimensional nature of any attempts for devising resource management plans in organisations, see Hax and Majluf (1983).

Management appreciates the possible link between resource allocation decisions and the assessment of performance. This awareness, however, is not reflected into management's actions. Schick (1990) argues that the use of performance measures in allocating resources is an old problem that is receiving renewed attention in recent years. Almost all methodologies of resource management discussed earlier seek to link resource allocation and performance measurement with a limited degree of success. This is undoubtedly a strong motivational factor of this thesis.

4. Resource management and improved performance

"The concept is simple- objectives, results, and resources should all be linked. The application is difficult[10]"

The ultimate objective to mould the resource management process into a "performance manifesto" constitutes a managerial aspiration. Undoubtedly, the financial/fiscal difficulties that many countries are facing reinforces the necessity for embracing performance in resource management decisions. This, although very prominent as an idea, has very many operational and practical difficulties. Schick (1990) argues that what is needed is merely to foster a managerial environment which is attentive to performance when funds are parcelled out. However, it seems unlikely that current resource allocation methods are able to encompass performance measurement effectively.

The numerous resource allocation processes discussed earlier on are not very successful in accommodating performance issues within the allocation process. This is mainly a result of the incompatibilities between the managerial processes adopted for allocating resources and those for assessing performance. Some of these incompatibilities are sought to be alleviated in the remainder of the thesis.

4.1. Relation between organisational structure and resource allocation

Heretofore, the discussion on the notion of resource allocation has been seen independently from the structural characteristics of MUOs.

- Does the organisational structure of MUOs have an impact on the adopted resource allocation processes ?
- Is resource allocation in the National Health Care systems, for instance, the same as the resource allocation in a grocery retailing company ?
- Are the same analytic methods needed for supporting resource allocation decisions made for covering running costs to the ones covering costs of investment projects ?

Resource allocation will be distinguished next into a number of categories depending on the structure and nature of the MUOs concerned.

---

[10] Auditor General of Canada, Annual report to the house of Commons (Ottawa: 1987), para 4.28.

Figure 1.3

Resource allocation & MUO structure

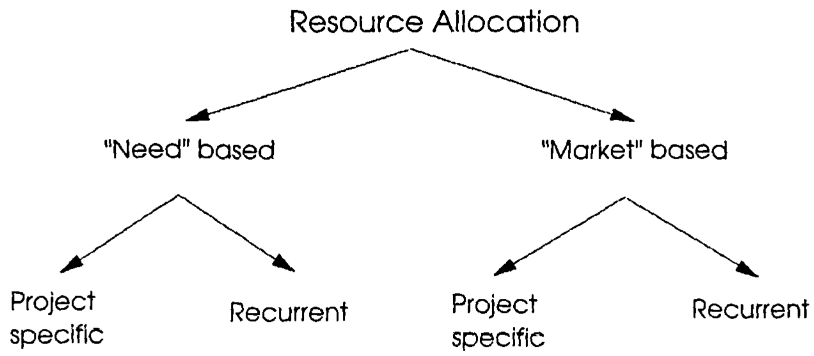


Figure 1.3, exhibits cases of resource allocation in relation to the nature of corresponding MUOs. In the first level of distinction a differentiation takes place based on the nature of the MUOs concerned. The second level distinguishes between the nature of resource allocation decisions within MUOs. The implications of the two level classification are discussed next.

4.1.1. "Need" versus "Market" resource allocation

As discussed earlier resources are allocated in not-for-profit MUOs in order to support their activities to meet the demand for services. Various resource allocation systems are used to estimate this demand, based on sociodemographic factors and then convert this "demand" into monetary equivalent terms. Examples[11] of "demand" driven systems are the Rate Support Grant (RSG) for the local authorities and the Resource Allocation Working Party (RAWP) formulae for the NHS in the UK. These "relative need" based systems often assume that the demand for services is more or less given whilst there is legislation guiding public MUOs to meet demand in the best possible way, (e.g. pupils allocated to schools). These systems typically allocate resources without, however, having any concern on how efficiently these resources are used. For example, the official document following the development of the RAWP (1976) states:

---

[11] In recent years the policy changes in British public policy have affected the purity of these "demand" driven systems by incorporating elements of efficiency into the allocation process. In the case of resource allocation within the District/Regional Health Authorities in the UK there has been a progressive shift towards an efficiency-based resource allocation system via the internal market processes.

"Resource allocation is concerned with the distribution of financial resources which are used for the provision of real resources. In this sense it is concerned with the means rather with the end. ***We have not regarded our remit as being concerned with how resources are deployed.***" (pp. 8 emphasis added).

Market based MUOs allocate resources to penetrate their markets. Therefore resources are allocated to support the operation of individual units in delivering goods/services in *market* environments. Although there is no legislation forcing market MUOs to provide services the ability to satisfy and increase demand determines their long run viability.

Satisfaction of demand for services is a key differential factor that affects resource allocation decisions. Not-for-profit MUOs seek to ***satisfy*** demand for services whilst profit making MUOs seek to ***attract and maintain*** demand from their markets. The development of the internal markets in selected areas of the public sector in the UK has elements of both the profit and non-profit characteristics. The demand driven characteristics of control and planning are treated with different methods, philosophy and priorities by the for-profit and not-for-profit organisations. These differences necessitated the need for considering the case of performance measurement and target setting in for-profit and not-for-profit organisations as separate cases.

4.1.2. Project specific versus recurrent resource allocation

MUOs' classification based on market and need based resource allocation can be taken one step further focusing on the actual use of resources. A distinction of resource allocation models can be made based on the differentiation between *project specific* and *recurrent* resource allocation.

Project specific resource allocation concerns capital investment decision making in MUOs. Capital investment decision making is made using investment appraisal techniques in profit and cost-benefit analysis in non-profit MUOs. **Project specific resources are allocated *a-priori* into different operating units, in the sense that, resources are allotted to support projects with limited life cycle and well defined purpose.**

A-priori resource allocation is associated with performance measurement by the very nature of the project selection process. Projects are selected on the basis of their higher expected performance (e.g. *returns on investment* or *net present value*) from alternative project proposals. However, the question of estimating the expected *performance* of candidate projects is not a trivial task. Investment management literature makes particular reference to the problems of estimating capital inflows and outflows of investment projects, Drury (1991). This includes forecasting and market research methods in the profit-making sector whilst the controversial *cost-benefit* analysis methods which are used in not-for-profit environments.

Allocation of recurrent grants seem to have a much "smoother" nature as a problem. Incremental resource allocation is a particularly attractive mechanism which has found wide application in these types of problems. Recurrent grants have general purpose and are used by individual operating units ***a-posteriori***. In other words there is no direct correspondence between allocated resources and tasks to be undertaken. This has profound implications on efficiency in cases where the allocation of resources is made incrementally based on past history. Public accountability is a notion that has been suggested for addressing the lack of direct evaluation of the actual use of allocated resources. Information concerning the degree of utilisation of resources allocated in the past would provide useful information about future resource allocation decisions.

Organisational structure has considerable influence on the way performance is measured and resources are allocated in MUOs. *Centralisation* and *Decentralisation* represent the generic form of two different organisational structures that are used to guide decision making in MUOs. The two systems differ on issues concerning information flow, hierarchies, decision making and autonomy among the different levels of MUOs' administration.

Performance measurement studies assess the appropriateness of actions taken by individual agents that have been assigned various types of contracts. As the links between agents and contracts are determined by the way MUOs are structured this needs to be reflected in the design of mechanisms for linking performance measurement and resource allocation. Chapters five and six will concentrate on the impacts of organisational structure in designing target-based resource allocation mechanisms in non-profit MUOs. Chapter eight on the other hand concentrates on developing methods that will assess performance customised to different levels of profit making MUOs.

5. Conclusions and research implications

This chapter focused on methodological issues of resource management in multi-unit organisations (MUOs). The linkage between managerial control and planning in MUOs is the main concern of this research. Control and planning will be represented by performance measurement and resource management respectively. It is generally accepted that the leading methodologies for allocating resources in for-profit and not-for-profit organisations have limitations for linking performance measurement and resource allocation decisions. A more rigorous framework for resource management and performance measurement was put forward to address the lack of integration between control and planning in MUOs.

The effects of MUOs' structure on the use of performance measurement and resource management methods were examined. This led to introducing a distinction between *"relative need"* and *"market based"* oriented decision making in MUOs management. A further distinction was also made, based on the actual purpose of allocated resources in MUOs. Project specific *(a-priori)* resource allocation is made to support predetermined projects in MUOs (e.g. capital investment) whilst recurrent grants *(a-posteriori)* are allocated into activity centres which spend resources at their discretion.

The immediate research implication from this analysis is that one needs to appreciate the structure of organisations and the nature of resource allocation decisions involved in the development of methods that link performance and planning mechanisms in MUOs. The problems of performance measurement and resource allocation are heavily influenced by political, strategic, managerial and other factors that are not always quantifiable. Therefore, the main contribution of the thesis should be on the development of **decision support** mechanisms and not **decision making** *per se*. Development of decision support mechanisms aim at progressing and improving organisational understanding about the importance of linking control and planning mechanisms in MUOs. This will also highlight the importance of improving and/or changing the way performance is assessed and resources are allocated in MUOs. Thus the real achievement will be to develop a sustainable framework of analysis of issues related with performance measurement and resource allocation in MUOs.

- END OF CHAPTER ONE -

Chapter 2

Recent developments for assessing performance: The evolution of frontier analysis methods

1. Introduction

During the last two decades, enormous attention has been given to the assessment and improvement of the performance of productive systems. Economic activities at the firm, industry, region or nation level are affected by the world-wide trend for improved performance. During these decades, national economies, in Japan for example, have gained economic advantage due to their ability to improve performance in their manufacturing and service delivery systems. On the other hand, the continuous economic recessions in the western world, the failure of the welfare state of the seventies in Europe and the subsequent failures of the liberal and neo-liberal systems of the eighties to control public spending and public deficit put enormous pressures on for-profit and not-for-profit organisations for the improvement of performance as a means to long run viability[1]. More recently, the collapse of the socio-economic structure in many eastern European countries has brought into focus the question of performance in a previously unknown scale. The question of improving performance has gained popularity among various political parties, and now, various initiatives can be found that discuss the issue of performance as a distinct political, economic and social concept.

[1] In the 1988-1993 recession in the UK most organisations sought to increase their productivity by cutting costs in order to survive under the adverse market conditions.

In the UK during the 1980's there erupted, in particular, a concern for *accountable management* within public sector organisations. Since then, a new generation of professionals and academics has flourished, with a greater emphasis upon the assessment of productive efficiency of systems. On the other hand, accountability and performance measurement in the private sector received increased attention, mainly spurred, by the unsuccessful experiences of auditing bodies (accounting firms) to uncover the true performance prospects of many profit making organisations, (e.g. see the Poly-Peck and BCCI fraud cases in the UK).

The revival of the performance measurement culture especially in the public sector but also in the private, has brought closer the previously unconnected disciplines that are, by nature, involved with the assessment of performance. Clearly, the assessment of performance has *political, economic, accounting and management science* dimensions which could be integrated for improving the way performance is assessed.

This chapter is organised as follows. Firstly, a review of performance measurement is made emphasising its multi-dimensional nature. It is then argued that individual disciplines can address the question of assessing performance in part. Thus a framework needs to be developed for integrating the strong features of different disciplines within a common performance measurement framework. This framework is called *frontier analysis* and its development includes stochastic and deterministic variants. This chapter concentrates on the nonparametric and deterministic aspects of *frontier analysis* with particular emphasis on issues that will be used later on in the thesis.

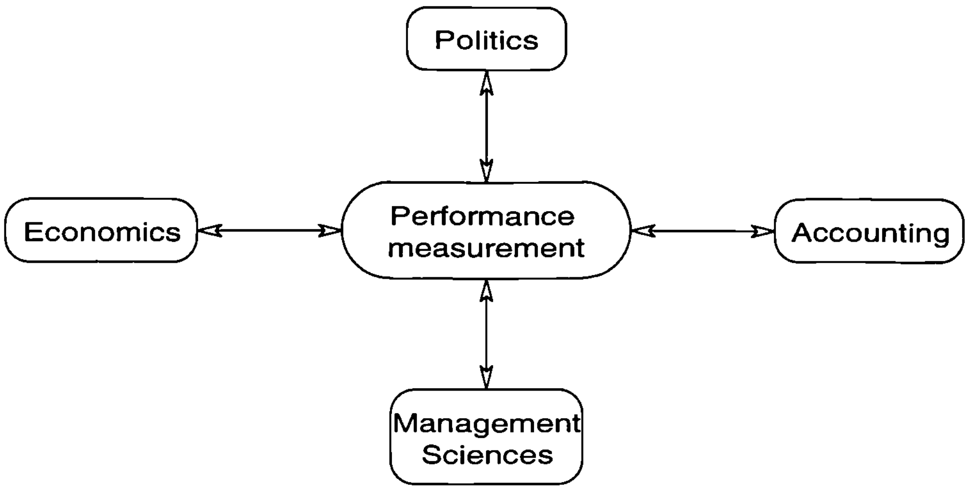
2. Aspects of the multi-disciplinary nature of performance measurement

For some authors the history of analysing performance of organisations is dated back to Plato's and Aristotle's discussions about the effectiveness of different military organisations, Hoagland (1964). Leonardo da Vinci in the fifteenth-century also studied performance questions concerning the labour effort in shovelling. Evidently, the concept of performance is an old problem in the history of sciences and philosophy. However, the glory of the systematic development and assessment of performance belongs almost exclusively in the post nineteenth century and in particular to F. Taylor who has been characterised as the father of scientific management[2]. Leaving outside the history about its originator, scientific management represents an attempt at improving the efficiency of various operating systems using laws and methods from the natural sciences.

Since the development of scientific management other related disciplines have advanced which define and consider the concept of performance from their own perspective. Figure 2.1, below lists four major disciplines with vast interests in the definition and assessment of performance.

Figure 2.1

Alternative views on performance measurement



Politics and performance

The assessment of the performance of systems has an inherent political dimension as it reflects the purpose and mission of the system. Based on a set of ideological principles political institutions seek to improve and enhance the performance of societies and economies. For many political scientists, the assessment of performance of institutions should primarily emphasise issues related with *freedom, access to power, decision making and rationalisation.* The latter conception shows that economic achievements in societies without democratic freedoms are not considered of any real value and it is argued that in the long run the performance of these systems will decline and eventually collapse. This is evident in the case of the ex-communist countries whose economic performance was impeded by the lack of real political democracy.

---

[2] Hoagland (1964), however, argues that many of Taylor's theories can be found published in previous research work.

Another area of linkage between politics and performance emanates from the concept of *political choice*. This can be demonstrated by using as an example the privatisation of the public utilities in the UK in the eighties. Advocates of privatisation argue that public control of telecommunications, electricity, water, etc. prevent these companies from operating efficiently, and therefore, should be run under private control. Competition and market conditions are expected to stimulate the economy, reduce costs, thereby, benefit the tax payer who will not have to contribute through taxation to the potential inefficient operation of companies like British Telecom. Performance can be used, therefore, to support decisions of political nature.

The opponents of privatisation, however, also use performance related arguments in order to object to privatisation. For example, while Mayston (1993) argues that the political decision to sell public assets has short run benefits on public finance he also claims that in the long run the public "loses" the opportunity to gain financially from the very high profits of these companies (e.g. £96 per second profits for British Telecom). The political debate on performance issues seems to be endless and undoubtedly has very subjective nature.

Accounting and performance

The concept of organisational accountability in both profit and not-for-profit organisations constitutes another dimension of performance. Accountability has strong political origins as it is the process that informs shareholders[3] on the propriety of decisions made in organisations. Historically, the accounting profession has been employed to generate information related to organisational performance. Booth and Cocks (1990) state that the accounting profession has traditionally been viewed as a neutral purveyor of the facts. Solomon's (1978) discussion of accounting methodology describes it as a cartographic method (accounting is financial mapmaking). Accounting's role in many types of institutions is growing with time. Critiques and advocates give various explanations for this phenomenon. The most important issue that arises from this debate, however, is whether accounting information provides sufficient evidence on the performance of organisations. Evidence obtained from the accounting literature emphasises that the current accounting practices give little assurance to shareholders on whether companies perform adequately.

---

[3] Shareholders in the public sector are assumed to be the taxpayers

An empirical study in the private sector by Citron and Taffler (1992) found no correlation between whether or not an audited firm received a going concern qualification and whether or not it did subsequently fail in the next 12 months. The more general case of creative accounting is well discussed in the accounting literature as an on-going and growing problem of the accounting profession, Griffiths (1992) and Smith (1992). Similar messages can also be found in the use of accounting practices in the public sector. For example, the demanding data requirements for implementing the Resource Management Initiatives (RMI) in the national health system have very high cost implications for developing appropriate information systems without substantial performance returns, (National Audit Office (1992).

Economics and performance

The economic approach to efficiency is perhaps the most elegant one. The reason for this lies in that economic phenomena, such as production, are axiomatically defined and then subsequently examined whether they are supported by real life facts. Productive efficiency is perceived in economics as the outcome of comparing the actual output of productive units against a theoretically defined maximum output given the resources they use. At the theoretical level, this is represented by the notion of the *production function* which, in short, represents an extreme relationship between inputs and outputs, and also, accounts for the maximum obtainable amount of output for a given level of input and *vice versa*.

The closest association between efficiency and economics can be found in the theory of production. Historically, one can find a number of key contributors that affected in one way or another by the development of what we shall introduce later as *frontier analysis*. Chambers (1990), in his monograph on production theory, uses the agricultural experiments (1820-1830) of Von Thuenen as a starting point of production theory whilst recognising Moore (1929) as one of the originators in using statistics to examine economic phenomena such as the marginal productivity theory.

According to Lovell (1993) a departure point of efficiency studies is Knight's (1933) work who defined efficiency as the ratio between outputs and inputs and furthermore discussed issues related with the selection of inputs/outputs for assessing efficiency. Chambers (1990) argues that, despite the earlier studies of production relationships, it was only after the seminal work by Cobb and Douglas (1928) that the estimation of production function became commonplace in economics. Since then, more flexible production function forms were developed and tested on empirical data.

Using theoretically defined production functions, economic theory progressed for a period of twenty-five years by examining the extent to which empirical data was compliant with the assumptions made in various production functions.

Lovell (1993) and Greene (1993) discuss in more detail the issues and problems related to the econometric methodology for assessing efficiency. One can argue that the estimation of econometric based frontiers, despite its advances during the last two decades, has yet to address the problem of selecting appropriate functional forms, the distributional problems of the inefficiency terms and the accommodation of multiple input-output cases. Schmidt (1985) and Thanassoulis (1993) discuss in more detail the pros and cons of econometric frontier estimation.

Management science and performance

As was mentioned earlier, the development of scientific management sought to borrow from the natural sciences for improving performance of socio-economic systems. In the post second world war period, scientific management was enhanced by "operational research" techniques. A large number of problems concerning resource allocation, location analysis, transportation planning, educational and health care planning and delivery were supported using tools like linear programming, project management (PERT, GANTT), decision trees, simulation, and queuing theory.

The main emphasis of OR methods was to provide decision support for planning. These efforts, however, did not consider the possibility of using operational research techniques in a control mission for assessing organisational performance. As Charnes and Cooper (1978) note:

> "Almost no attention (by operational research) has been devoted to improved procedures of accountability and/or other approaches to the control of management behaviour"

The original engagement of management science techniques, notably linear programming, in a pure control mode was launched by Charnes et al. (1978). This development initiated a whole new area of expansion for management sciences which, apart from its own development, has brought closer the previously diversified components of performance measurement.

In summary, performance measurement has very important political, accounting, economic and management science affiliations. The definition and assessment of performance measurement can vary from an abstract political concept to a set of performance indices reported by accounting auditors. There is a fundamental agreement that performance measurement needs to have a quantitative component at least where performance is assessed by some type of ordinal or nominal scale.

3. The evolution of frontier analysis

3.1. Background discussion

The previous discussion focused on the political, accounting, economic and management science dimension of performance measurement. As none of these disciplines can capture performance measurement in full, some synthesis towards a unified framework is necessary. An attempt towards that direction is made via frontier analysis.

Charnes *et al.* (1985) argue that Frich in (1935) analysing the chocolate production in France concluded that methods were needed to estimate frontier functions. Moreover, Aigner, Lovell and Schmidt (1977) mentioned:

> The theoretical definition of a production function expressing the maximum amount of output obtainable from given input bundles with fixed technology has been accepted for many decades. And for almost as long, econometricians have been estimating average production functions.
>
> [Aigner, Lovell and Schmidt (1977, p.21)]

Traditional economic approaches use theoretically justified production functions and test their behaviour on real data. Data sets that do not support the prespecified production functions lead to two possible conclusions -either the specification of the production function was inappropriate or the productive units in the analysis were very inefficient and therefore, could not give a sufficient statistical fit. **Unfortunately, problems like this do not seem to have any obvious answer.**

Farrell (1957) was the first to put forward an alternative, to the traditional economic, framework for assessing productive efficiency. Farrell suggested that productive efficiency should be assessed using empirical observations which avoids *a-priori* specification of functional forms. A pictorial representation of the work promoted by Farrell (1957) is given in Figure 2.2 below.

Figure 2.2

Farrell decomposition of efficiency

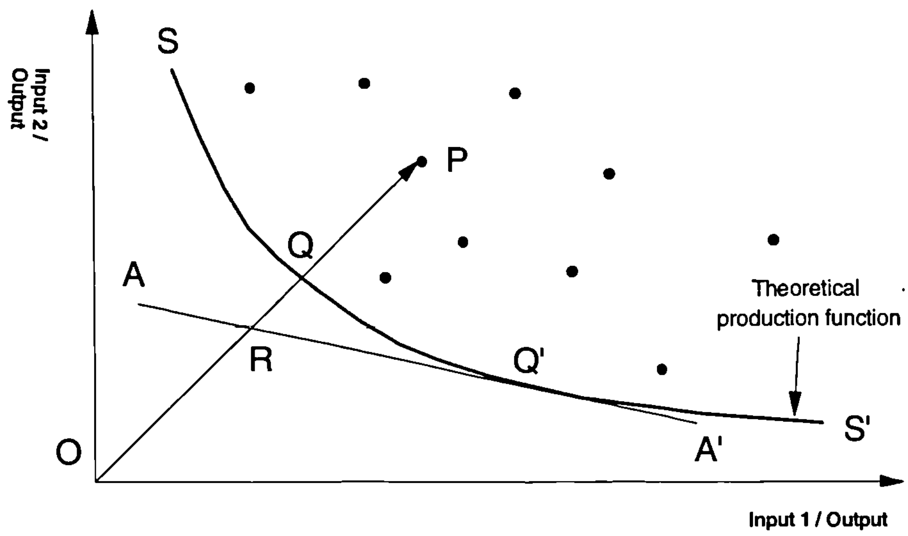


Figure 2.2 contains an example where decision making units (DMUs) require two inputs for producing one output. The input quantities have been standardised per unit of output produced and, therefore, the example has adopted a constant[4] returns to scale assumption. Suppose that the *efficient production function* is known and given by the curve SS'. In other words, the output that a perfectly efficient firm could obtain from any combination of inputs. Let us also assume that the prices p_1, p_2 for the two input quantities are also known, and the line AA' ($p_1x_1 + p_2x_2 = C$) has a slope equal to the ratio of the prices of the two inputs, where C is the cost of securing one unit of output and x_1, x_2 are quantities of input 1 and input 2 per unit of output produced.

Let us compare DMUs P and Q. They are both on the same ray from the origin which implies that they employ the same input mix (proportions). However, DMU Q produces the same output as P using only a fraction OQ/OP of the inputs used by Q. **We shall define, therefore, the ratio OQ/OP as the technical efficiency of DMU Q.**

It is equally important to find out the extent to which a firm uses the various factors of production in the best proportions in the light of their prices. Comparing points Q and Q' on the theoretical production function it is obvious that Q' uses the least cost input

[4] Increasing (decreasing) the inputs by a constant factor would increase (decrease) the outputs with the same factor.

combination per unit of output produced. The costs of production at Q' are a fraction OR/OQ of those at Q. **This ratio is defined as the price or allocative efficiency of Q and represents the price efficiency of all technically inefficient DMUs such as P that have been projected at point Q.**

Overall, if DMU P was technically and price efficient, its costs would be a fraction OR/OP of what they are. This ratio is called the overall efficiency of DMU P and can be decomposed into its technical and price efficiency components as follows: $OR/OP = OQ/OP \times OR/OQ$.

Farrell's work was innovative for a number of reasons:

- The need for specifying the functional form of production functions prior to estimating the productive efficiency from empirical data is relaxed,
- Efficiency was decomposed into *technical, allocative and overall* components. Later, he also added a *scale efficiency* component (Farrell and Feldstein (1962)),
- Linear programming in a performance measurement mode was used,
- The existence of multi-input and multi-output production functions was recognised without, however, providing a way of estimating them.

Farrell's work did not find an immediate widespread use and it is Aigner and Chu (1968) that launched the first attempt for assessing efficiency using Farrell's rationale. This approach was a linear programming based one and was extended later by Forsund and his associates (1976). The Aigner and Chu formulations will be discussed next.

Let us assume that a set of decision-making units (DMUs) use inputs $X \in \Re_{+}^{m}$ to produce a single output $y \in \Re_{+}$. It is also assumed that the amount of output produced using the available input quantities is given by the following function:

$$y_j = f(X_j;\beta) - u_j \qquad \text{(M2.1)}$$

Where β is a vector of parameters to be estimated, j the DMU and j=1,...,n is the number of DMUs. The disturbance term u_j is assumed to satisfy $u_j \geq 0$ and is deterministicaly defined.

Using the functional form introduced in M2.1 Ainger and Chu (1968) used the goal-programming model[5] in M2.2 to estimate the technical efficiency of DMUs.

$$\underset{\beta}{Min}\sum_{j=1}^{n}\left[f(X_j;\beta)-y_j\right] \qquad \text{(M2.2)}$$

$$s.t.\left[f(X_j;\beta)-y_j\right]\geq 0 \quad \forall j$$
$$\beta\geq 0$$

This formulation allowed Aigner and Chu (1968) to estimate deterministic parameters β that described the structure of the production function.

The efficiency of individual DMUs E_j^{A-C} was obtained as: $E_j^{A-C}={y_j}\big/{f(X_j;\beta^*)}$.

This type of frontier was used mainly for assessing the <u>industry</u> production function or the so-called structural efficiency. The notion of the industry production function is elaborated further in chapter four when the estimation of global organisational targets is discussed.

Along side of all these developments, a parallel stream of economic thought was developed by Leibenstein (1976) postulating the existence of nonallocative inefficiency in production (i.e. non-optimal mix of inputs in terms of cost minimisation or profit maximisation). Frantz (1992) argues that until that time economists thought mainly about allocative inefficiency and assumed that firms were always maximising their technical efficiency due to the market's pressure. X-efficiency as a concept is more ambitious than the technical efficiency as defined by Farrell (1957). Leibenstein (1976, 1987) clarified these distinctions:

> "X-efficiency is not the same thing as what is frequently referred to as technical efficiency, since X-efficiency may arise for reasons outside the knowledge or capability of management attempting to do the managing ... . In other words, it is not only a matter of techniques of management, or anything else technical in carrying out decisions that is involved in X-efficiency",
>
> [Leibenstein, 1976, p.27]

Lovell (1993) argues that there is scope for linking the literature of X-efficiency with the performance measurement literature which was evolved in the post Farrell (1957) period.

---

[5] In their formulation Aigner and Chu (1968) have taken a logarithmic transformation of inputs and outputs as they assumed a so-called Cobb-Douglas production function linking inputs to the single output produced ($y_j=c\prod_i\left(x_{ij}\right)^{\beta_i}$).

Leibenstein and Maital (1992) seem to agree with this as they appreciate the potential similarities between X-efficiency and frontier analysis.

The turning point after Farrell (1957) in the assessment of productive efficiency came via two parallel attempts[6] for assessing performance at the firm level. The economic attempt was made by Fare and Lovell (1978), whilst the operational research attempt was made by Charnes, Cooper and Rhodes (1978). A new "technique" called *data envelopment analysis* emerged from these initial attempts opening a very wide research area which since then has gained widespread development.

> In this research *frontier analysis* denotes the methodology that draws upon *economic, engineering economic, econometric, system, accounting and operational research* disciplines with the objective to assess productive efficiency based on observed behaviour using the minimal possible sets of assumptions.

The remaining sections of this chapter attempt to review the technical side of frontier analysis for assessing productive efficiency. The models presented refer to efficiency models in the post Charnes and Cooper (1978) period.

3.2. Frontier analysis via Data Envelopment Analysis

This section focuses on the development of a framework for assessing the technical efficiency of DMUs using the principles of frontier analysis. Frontier analysis is perceived as having three interrelated components.

- A systems' component
- A mathematical programming component
- A decision support component

It is argued that in the last fifteen years the development of frontier analysis has been drifting towards the mathematical programming component without a balanced development of the other two components. This thesis will attempt to emphasise the remaining two components.

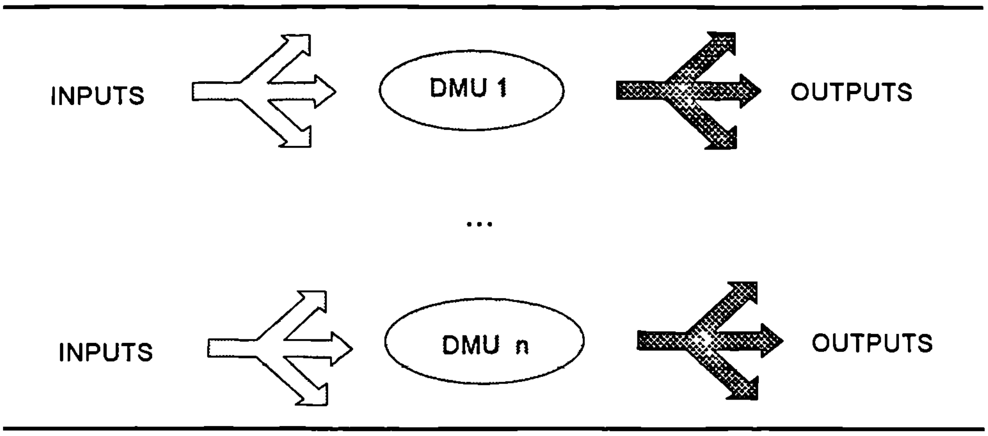
Frontier analysis seeks to investigate the performance of productive systems which employ input factors to deliver outcomes as represented diagrammatically in Figure 2.3.

---

[6] These attempts are limited to the non-parametric deterministic methodologies. One needs to note that the presence of the stochastic methodologies for assessing productive efficiency (see Jondrow et al. (1982)) which are outside the scope of the current thesis.

Figure 2.3

Input-output systems for frontier analysis



The very nature of performance measurement is heavily influenced by the inputs/outputs identified in a production process. For example, assessing the performance of schools using as inputs the resources available at a school (no. of teachers, facilities and expenditure) and as an output the examination achievements of pupils one can examine the rate that individual schools utilise their resources by achieving high examination results. If, however, the input list of the assessment included information on the entry standards, as well as sociodemographic background of pupils the assessment would yield information concerning the value added at schools, Thanassoulis and Dunstan (1994). Apart of the nature of the input/output used in assessing performance, questions can also be raised concerning the appropriate number of inputs/outputs for describing an activity process.

In economic literature one can find extreme opinions about the role of input-output systems in assessing performance. For instance, Knight (1933) has argued that if all inputs and outputs are included in assessing the efficiency of DMUs then they will all get an efficiency of unity (100%). Knight, therefore, suggests for redefining productivity using only the "useful" inputs and outputs. More recently Stigler (1976) and Ray (1988) have argued that the measured inefficiency may reflect the failure to incorporate the right variables and the right constraints and to specify the right economic objective, of the production unit.

The definition of input-output models incorporates *behavioural* aspects of an organisation with profound implications for the subsequent assessment of performance. The frontier analysis literature is short of systematic attempts for developing input-output models for assessing efficiency. The methods chiefly used are statistical significance, influence diagrams and expert opinion. As Varian (1988) argues, statistical significance has been given emphasis in setting up input-output models without, however, supporting always the *economic* significance of the included variables. The process for defining input-output production sets can be enhanced linking frontier analysis with other methodologies like the *cognitive mapping* and *systems dynamics* approaches, Eden (1988).

3.3. Production possibilities and efficient frontiers

The discussion of any production activity in economic theory must draw on the notion of the *production possibility set.* The production possibility set Φ is, in theory, an unknown and it will therefore either be defined in abstract or it will be defined using observed production units. Let us suppose that we have data on a set of j = 1,...,n DMUs and each DMU use inputs $X \in \Re_{+}^{m}$ to produce outputs $Y \in \Re_{+}^{s}$. Thus, DMU j uses amount x_{ij} of input i to produce amount y_{rj} of output r. A *referent production set* (or production possibility set) contains all input-output feasible combinations. Formally this can be stated as follows:

$$\Phi \equiv \{(X,Y) | \text{Input vector X can produce output vector Y}\} \qquad \text{(M2.3)}$$

The definition of the production possibility set is strengthened further using the following postulates:

***Postulate* 1.** (*Non-Stochastic*) All observed operating DMUs are included in the referent set,

***Postulate* 2.** (*Inefficiency or Free disposal*)

$$(a) \text{ If } (X,Y) \in \Phi \text{ and } X' \geq X, \text{then } (X',Y) \in \Phi$$
$$(b) \text{ If } (X,Y) \in \Phi \text{ and } Y' \leq Y, \text{then } (X,Y') \in \Phi$$

Postulate 3. (*Ray Unboundedness*)

$$\text{If } (X,Y) \in \Phi \text{ then } (kX, kY) \in \Phi \ \forall k > 0$$

Postulate 4. (*Convexity)*

$$\text{If } \left(X^j, Y^j\right) \in \Phi, j = 1,\ldots,n \text{ and } \mu_j \text{ are non-negative indices}$$
$$\text{such that } \sum_{j=1}^{n} \mu_j = 1, \text{ then } \left(\sum_{j=1}^{n} \mu_j X^j, \sum_{j=1}^{n} \mu_j Y^j\right) \in \Phi.$$

Postulate 5. (*Minimality assumption*)

Φ is the intersection of all $\hat{\Phi}$ satisfying Postulates 1, 2, 3, 4 and subject to the condition that each of the observed vectors $\left(X^j, Y^j\right) \in \hat{\Phi}, j = 1,\ldots,n$.

Postulates 1, 2, 3, 4 and 5 can be used to define a constant returns to scale (CRS) production possibility set shown below in M2.4.

$$\Phi_{CRS} \equiv \left\{(X,Y) \in \Re_{+}^{m+s} \middle| X \geq \sum_{j=1}^{n} \lambda_j X^j, Y \leq \sum_{j=1}^{n} \lambda_j Y^j, \lambda_j \geq 0\right\} \qquad \text{(M2.4)}$$

Exclusion of postulate 3 will lead to the definition of a variable returns production possibility set (VRS) shown below in M2.5.

$$\Phi_{VRS} \equiv \left\{ \begin{array}{l|r} (X,Y) \in \Re_+^{m+s} & X \geq \sum_{j=1}^{n} \mu_j X^j, Y \leq \sum_{j=1}^{n} \mu_j Y^j \\ & \mu_j \geq 0 \text{ are scalars with } \sum_{j=1}^{n} \mu_j = 1 \end{array} \right\} \qquad \text{(M2.5)}$$

The CRS and VRS production possibility sets in M2.4 and M2.5 have in common a fundamental feature that includes as members of the production possibility set *linear combinations* of inputs and outputs of observed DMUs. The convexity property, however, affects the frontier of the production possibility set. In the case of CRS the frontier is defined as a *conical hull* whilst in the case of VRS the frontier is defined as a *convex hull* of the production possibility set.

To each production possibility set there corresponds an efficient frontier which consists of a subset of its DMUs that satisfy the property of efficiency. Notice here that the concept of an efficient frontier is linked with the production possibility set. Technical efficiency can be defined as input-saving, output-augmenting or a combination of the two. In an input-saving sense the efficiency E_j^F of DMU j under constant returns to scale can be defined as follows:

$$E_j^F = \min\left\{\theta \middle| X \leq \theta X^j, Y \geq Y^j \; and \; (X,Y) \in \Phi_{CRS}\right\} \qquad \text{(M2.6).}$$

All production possibilities (X_j, Y_j) with a $\theta = 1$ are called Farrell-efficient DMUs and constitute the Farrell efficient frontier. This definition, however, is not sufficient for defining "truly" efficient frontiers in the *Pareto* sense. Koopmans (1951) defined technical efficiency as follows:

> A producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a decrease in at least one output. [Koopmans, 1951, pp. 60]

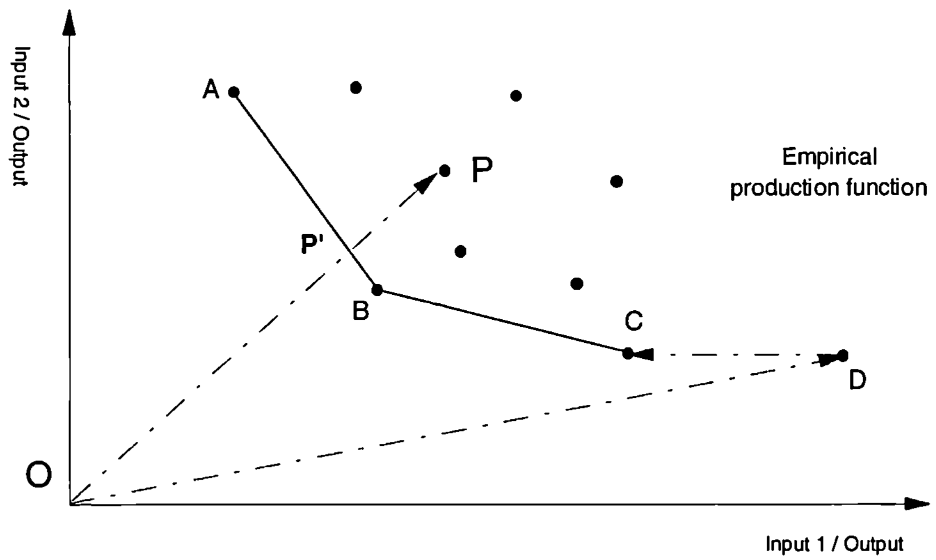
The mathematical expression of this definition is as follows:

$$E_j^{P-K} = \max\left\{ s+d \left| \begin{array}{l} X \leq X^j - s, Y \geq Y^j + d \; and \; (X,Y) \in \Phi_{CRS} \\ s \in \Re_+^m, d \in \Re_+^s \end{array} \right. \right\} \qquad \text{(M2.7)}$$

An optimal solution of $s^*+d^* = 0$ to model M2.7 indicates that the corresponding assessed DMU j is a *Pareto-Koopmans* efficient DMU. DMUs satisfying this criterion constitute an efficient subset of the frontier of a production possibility set. (In the remainder of the thesis the term efficient frontier will always correspond to DMUs that satisfy the *Pareto-Koopmans* criterion).

Farrell's efficiency in M2.6 is based on the radial contraction factor θ which does imply that at the boundary for some individual inputs (outputs) there is no scope for further reduction (expansion). Koopmans efficiency in M2.7 investigates the performance of each input and output of assessed DMUs beyond the radial contraction factor θ. Let us illustrate the distinction between Farrell and Koopmans frontiers graphically using a two input production possibility set standardised per unit of output as exhibited in Figure 2.4.

Figure 2.4
Farrell Vs. Pareto efficient frontiers



In Figure 2.4 above it is assumed that the theoretical production function is an unknown and therefore efficiency is estimated on the basis of an empirical production function based on the performance of DMUs A, B and C. The technical efficiency of DMU P is defined as the ratio OP'/OP and it estimates the proportionate excess use of input 1 and input 2 in producing one unit of output. In the case of DMU D (and each DMU on its horizontal expansion) the Farrell test will give an efficiency equal to one as the OD ray from the origin meets DMU D without any interference from the efficient frontier. Is DMU D, then an efficient one ? Clearly not, as DMU C uses that same amount of input 2 and less amount of input 1 to produce one unit of output. Farrell in his work in (1957) appreciated the problem caused by these type of DMUs which he called "DMUs at infinity" without, however, providing any methods for identifying the true efficiency of these DMUs. Using the Koopmans definition of efficiency it is clear the DMU D is an inefficient DMU.

3.4. Linear programming models for assessing efficiency

The frontier analysis discussion has so far succeeded in providing a systematic definition of production possibility sets and their efficient frontier. The next step will be to define some type of "metric" that would enable us to project inefficient DMUs on the efficient frontier of their production possibility set in M2.4. This can be done using the linear programming models developed by Charnes, Cooper and Rhodes (1978) which operationalised and extended the earlier work by Farrell (1957).

The technical efficiency of a DMU j_o can be obtained in using the two-stage linear programming model in M2.8. The assessment of efficiency can be done by using an input contraction (CRS^I) or output expansion (CRS^O) orientation. For convenience sake, only the first stage of this process will be stated in the remainder of the thesis, assuming however, that any numerical calculations for assessing efficiency require this two stage process.

Linear programming for assessing technical efficiency (M2.8)

CRS^I - Input contraction	
Stage 1 - Contraction factor (θ)	Stage 2 - Pareto test (s_i,s_r)
$\underset{\theta,\lambda'_j}{Min}\ \theta^* = \theta$	$\underset{s_i,s_r,\lambda_j}{Min}\ -\sum_{i=1}^{s} s_i^- - \sum_{r=1}^{m} s_r^+$
$\sum_{j=1}^{n} \lambda'_j x_{ij} \le \theta x_{ij_o} \quad \forall i$	$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta^* x_{ij_o} \quad \forall i$
$\sum_{j=1}^{n} \lambda'_j y_{rj} \ge y_{rj_o} \quad \forall r$	$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{rj_o} \quad \forall r$
$\lambda'_j \ge 0, \theta$ free	$s_i^-, s_r^+, \lambda_j \ge 0$
CRS^O - Output expansion	
Stage 1 - Expansion factor (z)	Stage 2 - Pareto test (d_i,d_r)
$\underset{z,\lambda'_j}{Max}\ z^* = z$	$\underset{d_i^-,d_r^+,\lambda_j}{Min}\ -\sum_{i=1}^{s} d_i^- - \sum_{r=1}^{m} d_r^+$
$\sum_{j=1}^{n} \lambda'_j x_{ij} \le x_{ij_o} \quad \forall i$	$\sum_{j=1}^{n} \lambda_j x_{ij} + d_i^- = x_{ij_o} \quad \forall i$
$\sum_{j=1}^{n} \lambda'_j y_{rj} \ge z y_{rj_o} \quad \forall r$	$\sum_{j=1}^{n} \lambda_j y_{rj} - d_r^+ = z^* y_{rj_o} \quad \forall r$
$\lambda'_j \ge 0, z$ free	$d_i^-, d_r^+, \lambda_j \ge 0$

Where x_{ij} is the level of i^{th} input and y_{rj} is the level of r^{th} output of the j^{th} DMU; m and s are the dimensions of the input and output space respectively and n is the number of DMUs.

The solution process in M2.8 yields input contraction or output expansion efficiencies obtained from a two-stage process. Stage 1 seeks to identify the maximum *pro rata* input decrease or output increase. The optimal solutions obtained correspond to the *Farrell* type of efficiency discussed earlier on. Stage 2 investigates the potential extra input reduction or output expansion beyond that which is already achieved at the first stage[7]. The combined solutions from stage 1 and stage 2 can be used for identifying *Pareto-Koopmans* efficient DMUs.

A DMU j_O will be efficient if and only if the solution of the input contraction LP model is $\theta^* = 1$ and $s_i^{-*} = s_r^{+*} = 0 \; \forall i, r$. Similarly DMU j_O will be efficient if and only if the solution of the output expansion LP is $z^* = 1$ and $s_i^{-*} = s_r^{+*} = 0 \; \forall i, r$.

Charnes *et al.* (1978) baptised the method used for assessing efficiency, **"data envelopment analysis" (DEA),** in an attempt to describe the rationale of the method: use the relatively efficient DMUs of a production possibility set in order to create an *efficient envelope* for inefficient DMUs. From the solution of model M2.8 emanate a number of observations summarised below:

- Model M2.8 assesses the efficiency of DMUs under constant returns to scale by solving an LP problem for each observed DMU in the production possibility set Φ_{CRS}.

- For each assessed DMU j_O the solution process of M2.8 seeks to identify a comparator (or combination of) efficient DMUs $\left(\sum_{j=1}^{n} \lambda_j^* x_{ij}, i = 1,\ldots,m; \sum_{j=1}^{n} \lambda_j^* y_{rj}, r = 1,\ldots,s\right)$ that dominate j_O in all input/output dimensions. It can be argued that this mechanism is based on an *offensive* behaviour of the relatively efficient DMUs over the inefficient ones.

- The technical efficiency of a DMU j_O is measured by the radial contraction θ^* or radial expansion factor z^* respectively. Despite the suggestions for adjusting the radial efficiencies to incorporate the slacks, Ali (1991), it is argued here that efficiency should be defined from the radial factors obtained in the first stage of M2.8.

- There is an inverse relationship between the input contraction θ^* and output expansion z^* efficiencies under constant returns to scale. It can be shown (see Seiford and Thrall (1990)) that in the optimal solution of M2.8 the following

---

[7] The original formulation by Charnes *et al.* (1979), compounds the two stages in one by including the slack variables in the objective function of the first stage multiplied by very small coefficients. This method despite its simplicity creates computational difficulties by Ali and Seiford (1989).

relation holds: $\theta^* = \frac{1}{z^*}$. This, however, should be seen as a special case (constant returns to scale) and not as a general rule.

The linear programming model in M2.8 will be used in the rest of the thesis as the basic component for assessing the efficiency of decision making units.

3.5. An alternative formulation of DEA: The defensive LP model

The mathematical programming models employed in M2.8 were interpreted as "offensive" DEA models due to the use of composite DMUs $\left(\sum_j \lambda_j^* X_j, \sum_j \lambda_j^* Y_j,\right)$ as comparators to inefficient DMUs. An alternative (value based) formulation can, however, be given for assessing efficiency based on the dual form of the models in M2.8. Value based models will present DEA in the light of a generalised total factor productivity index, often met in accounting and economic literature.

A total factor productivity index TFP of a DMU or firm can be defined as the weighted sum of its outputs divided by the weighted sum of its inputs. Using the notation used earlier in DEA the TFP_j of DMU j is given in M2.9.

$$TFP_j = \frac{\sum_{r=1}^{s} u_r' y_{rj}}{\sum_{i=1}^{m} v_i' x_{ij}}; \quad u_r', v_i' \geq 0. \tag{M2.9}$$

The selection of the weights for inputs v_i' and outputs u_r' respectively in M2.9 leads to the value of TFP for individual firms. The absence of market prices that will convert TFP into monetary terms necessitates the assignment of arbitrary weights reflecting the relative importance of individual inputs/outputs in assessing productivity. Very often an assignment of equal weights among inputs and outputs is used to resolve the problem of input/output aggregation. The presence of non-commensurate inputs/outputs causes extra difficulties in the assessment of weights of importance.

In DEA the TFP formulation is enhanced by selecting weights based on a comparative basis. In other words, the weights are treated as variables of an optimisation problem that seeks to maximise the TFP of individual DMUs, subject to the constraint that no other DMU can achieve a TFP value higher than unity (or some other upper limit). This is similar to the engineering definition of efficiency where the energy produced by a process cannot exceed the energy consumed for its generation, Charnes *et al.* (1985d).

The mathematical formulation of this model is given in M2.10 as was developed by Charnes *et al.* (1978) and modified by Charnes *et al.* (1979).

$$
\begin{aligned}
&\underset{v_i', u_r'}{Max} \quad \frac{\sum_{r=1}^{s} u_r' y_{rj_o}}{\sum_{i=1}^{m} v_i' x_{ij_o}} \\
&\text{s.t.} \quad \frac{\sum_{r=1}^{s} u_r' y_{rj}}{\sum_{i=1}^{m} v_i' x_{ij}} \le 1 \quad \forall j = 1, \ldots, n \\
&\qquad \frac{u_r'}{\sum_{i=1}^{m} v_i' x_{ij_o}} > 0 \quad \forall r \\
&\qquad \frac{v_i'}{\sum_{i=1}^{m} v_i' x_{ij_o}} > 0 \quad \forall i
\end{aligned}
\qquad \text{(M2.10)}
$$

The model in M2.10 is a linear fractional programming problem which can be converted to an ordinary linear programme using the Charnes and Cooper (1961) transformation. However, the important feature of this model lies more on the interpretation of its mechanism rather than on its mathematical transformation.

> An assessed DMU j_O "chooses" the set of weights $(v_i'^{j_o}, u_r'^{j_o})$ that maximise its efficiency TFP_{j_o}. The same weights are then attached to all other DMUs which try to "defend" their efficiency. If no other DMU reaches a higher efficiency score using the weights of the assessed DMU j_O the DMU is efficient; otherwise inefficient.

Based on this rationale the model in M2.10 will be called a *defensive* DEA model . The linear programming equivalent (for the output expansion case[8]) of model M2.10 is provided next in M2.11 which is the dual mathematical model for the output expansion in M2.8.

---

[8] An equivalent formulation holds for the input contraction case which was omitted to avoid repetitions.

Offensive - Defensive output expansion DEA models

CRS^O - Offensive	CRS^O - Defensive **(M2.11)**
$\max\limits_{\lambda_j, z} \quad z$ $-\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = -x_{ij_o} \quad \mathrm{i} = 1,\ldots,\mathrm{m}$ $\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = z y_{rj_o} \quad \mathrm{r} = 1,\ldots,\mathrm{s}$ $z \; free$ and $\lambda_j \geq 0, \forall j \;\; s_i^-, s_r^+ \geq 0$	$\min\limits_{v_i, u_r} \quad E_{CRS}^O = \sum_{i=1}^{m} v_i x_{ij_o}$ $s.t. \qquad \sum_{r=1}^{s} u_r y_{rj_o} = 1$ $\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} - t_j = 0$ $v_i, u_r, t_j \geq 0$

Strong Complementary Slackness Condition (M2.12)
$\lambda_j^* \times t_j^* = 0 \qquad \lambda_j^* + t_j^* > 0$ $s_i^{-*} \times v_i^* = 0 \quad and \quad s_i^{-*} + v_i^* > 0.$ $s_r^{+*} \times u_r^* = 0 \qquad s_r^{+*} + u_r^* > 0$

The solutions obtained via the offensive and defensive DEA models are linked via the duality theorem in mathematical programming. They, therefore, yield the same objective function value whilst their variables are linked via the Strong Complementary Slackness Condition (SCSC) stated in M2.12. The importance of the complementarity between the solutions of the two problems will be discussed more extensively in the fourth chapter of this thesis.

3.6. Decomposing technical efficiency

The technical efficiency obtained by M2.8 is under constant returns to scale. Banker *et al.* (1984) relaxed this assumption and developed ways for disentangling technical efficiency into *scale* and *pure technical components.* The idea is illustrated using the small numerical example in Figure 2.5.

Figure 2.5

Efficient frontiers & returns to scale assumptions

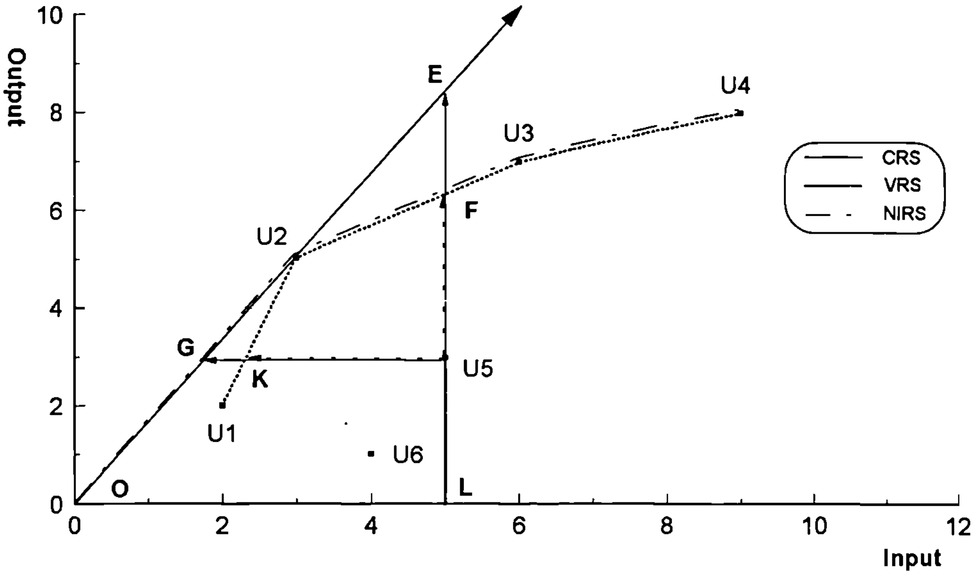


Figure 2.5 represents a single input-output production technology made of DMUs U1-U6. Under the assumption of constant returns to scale, DMU U2 is the only efficient DMU as it has the highest output per unit of input (5/3). The efficient frontier is in this case made of the conical hull OU2E which is an envelopment surface that can be stated as $\{\lambda(X_2, Y_2)|\lambda \geq 0\}$, where X_2 and Y_2 are the inputs/outputs of DMU U2. DMU U5 should, therefore, either expand its output along the ray U5E or contract its input along the ray U5G in order to be technically efficient.

The frontier is developed under the assumption that DMU U2 can be extrapolated to points, say, E and G without altering its output to input ratio. Relaxing this assumption one may redefine the efficient frontier without allowing scale extrapolations. The best observed practices, therefore, will be selected on the basis of performance given their scale of operation. The frontier in Figure 2.5 will be redefined, therefore, to be the piece-wise segment U1U2U3U4. This frontier is a variable returns to scale (VRS) frontier and is made of convex combinations of the extreme points lying on its surface.

DMU U5 is an output-inefficient DMU projected on the envelopment surface U2U3 defined as $\{\mu_2(X_2, Y_2) + \mu_3(X_3, Y_3)|\mu_2 + \mu_3 = 1\}$. DMU U5, in this case, should expand its output by a factor of 2.10 which is equivalent with the ratio $LF/_{LU5}$ (47.5% output efficiency). In the input side DMU U5 should contract its input by a factor of 2.155 which corresponds to the segment U5K (46.4% efficiency).

There are a number of important observations emanating from the VRS frontier.

- The orientation of the efficiency assessment (input or output) affects the facet of the projection and therefore input and output efficiency of a DMU will not be the same.
- Combining the constant and variable returns to scale frontier we can define a new efficiency component, namely the scale efficiency of a DMU. For example the output scale efficiency of DMU U5 is LF / LE .

Finally a frontier of mixed character can be developed where extrapolations are permitted for only a subset of efficient DMUs. Let us consider the piecewise segment OU2U3U4. This will be defined as a non-increasing returns to scale (NIRS) frontier. Under this assumption, the scale size of technical efficient DMUs, e.g. DMU U2, can be extrapolated for comparisons with smaller, e.g. DMU U1, but not larger DMUs, e.g. DMU U3. This type of frontier is used very rarely in the DEA literature, Tulkens *et al.* (1993), Fare *et al.* (1983).

Banker *et al.* (1984) and Fare *et al.* (1985) extended the original DEA models in order to estimate efficiency under the new set of assumptions. The offensive and defensive (dual) version of these models for an output expansion case are provided in M2.13.

DEA models for pure technical efficiency (M2.13)

Offensive output expansion variable returns to scale VRS	**Defensive output expansion variable returns to scale VRS**
$\underset{\mu_j,\rho}{\text{Max}}\ \rho$ $\sum_{j=1}^{n}\mu_j x_{ij} \le x_{ij_o}\quad i=1,\dots,m$ $-\sum_{j=1}^{n}\mu_j y_{rj} \le -\rho y_{rj_o}\quad r=1,\dots,s$ $\sum_{j=1}^{n}\mu_j = 1$ ρ *free* and $\mu_j \ge 0, \forall j$	$\underset{v_i,u_r}{Min}\ E^o_{VRS} = \sum_{i=1}^{m} v_i x_{ij_o} + \omega$ $s.t.\quad \sum_{r=1}^{s} u_r y_{rj_o} = 1$ $\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} + \omega \ge 0$ $v_i, u_r \ge 0; \omega$ free
Offensive output expansion non-increasing returns to scale NIRS	**Defensive output expansion non-increasing returns to scale NIRS**
In previous formulation change $\sum_{j=1}^{n}\mu_j \le 1$	In above formulation change $\omega \ge 0$

Model M2.13 differs from the original DEA model in M2.8 in that it has an extra (convexity) constraint added in the offensive model and the extra free variable (ω) added in the defensive model. The changes for the non-increasing returns to scale are also provided in the last row of the formulation. The solution of model M2.13 yields estimates on the pure managerial efficiency ($1/\rho^*$)of assessed DMUs. Combining the technical efficiency from M2.11 and pure technical efficiency from M2.13 we can obtain the output expansion scale efficiency index as the ratio z^* / ρ^*.

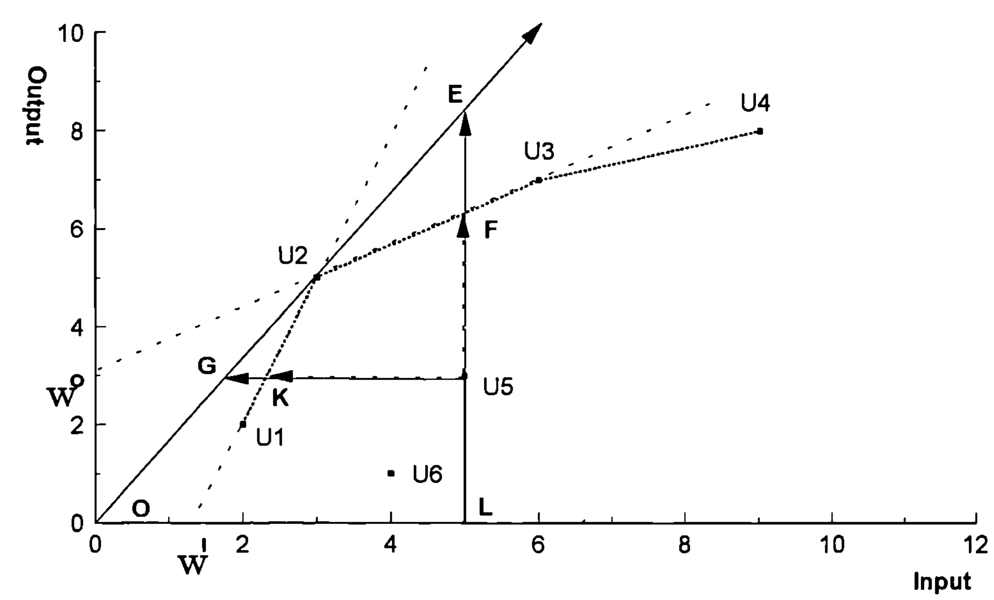
3.7. Economies of scale & most productive scale size

The definition of DEA efficient frontiers has been associated with scale related issues. As a result efficient frontiers that satisfy three different assumptions of returns to scale were developed. A constant returns to scale frontier assumes that proportionate inputs' reductions (increases) would be followed by equiproportionate outputs decreases (increases). A variable returns to scale assumption allows deviations in both directions. These directions constitute the nature of scale inefficiency and are listed below.

- A DMU operates under local increasing returns to scale if a proportionate increase (decrease) to its inputs will result in a higher than proportionate increase (decrease) to its outputs.
- A DMU operates under local decreasing returns to scale if a proportionate increase (decrease) to its inputs will result in a lower than proportionate increase (decrease) to its outputs.

The discussion will be facilitated using the geometric illustration of Figure 2.6.

Figure 2.6
Classifying local economies of scale



As discussed earlier, DMU U2 has the highest ratio of output per DMU of input and is, therefore, the only efficient DMU under an assumption of constant returns to scale. A different efficient frontier is, however, obtained under a variable returns to scale assumption. DMUs U1 and U3 and U4 are, therefore, pure managerially efficient but scale inefficient DMUs.

Banker *et al.* (1984) observed that the point of intersection between the constant and variable returns to scale frontiers can be used for characterising the nature of scale inefficiencies for individual DMUs. The segments below point U2 such as U1U2 characterise local increasing returns to scale whilst the segments above point U2 such as U2U3 and U3U4 characterise local decreasing returns to scale. An immediate implication from this definition is that DMUs will get returns to scale characterisation depending on the segment of the VRS frontier they are projected.

A numerical criterion for characterising increasing or decreasing returns to scale can be obtained from the extent to which CRS efficient DMUs adjust their scale size to be compared with inefficient DMUs. In the case of DMU U5, its comparator (DMU U2) needs to expand its performance up to point E. This can be expressed using the scale indicator
Λ^{O}=OE/OU2. As $\Lambda^{O} > 1$ this implies that the non-optimal scale of DMU U5 is larger than the scale of DMU U2 which operates under constant returns to scale. For the input contraction orientation of the efficiency of DMU U5 its comparator (DMU U2) needs to move at point G. This can be expressed using the scale indicator Λ^{I}=OG/OU2 which, as it is less than unity, indicates that the non-optimal scale of U5 is smaller that DMU's U2 and therefore DMU U2 operates under local increasing returns to scale.

The notion of the Most Productive Scale Size (MPSS) of scale inefficient DMUs is also related with the assessment of scale inefficiencies in DEA. Banker (1985) defined MPSS for an input-output combination as follows.

- A production possibility $(X,Y) \in \Phi_{VRS}$ is a MPSS for its input and output mix, if and only if for all $(\alpha X, \beta Y) \in \Phi_{VRS}$ we have $\alpha \geq \beta$.

Banker and Thrall (1992) showed that a MPSS corresponds to points of the efficient frontier that maximise the average productivity (= α/β) of a production possibility given its input output mix $(\alpha X, \beta Y)$. According to that definition DMU U2 in Figure 2.6 is the only MPSS DMU. This implies that the projection of U5 on points G or E albeit results in efficient positions they do not maximise the average productivity of DMU U5. Banker *et al.* (1984) and Banker and Thrall (1992) have shown that the scale factor Λ, as defined above, indicates the extent to which DMUs operate away from their most productive scale size.

The MPSS for DMU U5, therefore, can be estimated as $\left({X_E}/{\Lambda^O}, {Y_E}/{\Lambda^O}\right)$ for output expansion or $\left({X_G}/{\Lambda^I}, {Y_G}/{\Lambda^I}\right)$ for input contraction strategies. **In the two dimension case these two points coincide, without this being the case in the multiple input/output case.**

To characterise economies of scale in a multi-input multi-output case Banker *et al.* (1984) gave a set of criteria which were generalised latter by Banker and Thrall (1992). A different set of criteria has also been suggested by Fare *et al.* (1985) and Forsund *et al.* (1992). In our view, the question of identifying economies of scale using DEA will need further elaboration for comparing and integrating the alternative tests suggested in the literature. Table 2.1 lists the Banker and Thrall (1992) criteria for characterising returns to scale.

Table 2.1

Criteria for identifying returns to scale (output expansion case)

Characterisation	Offensive model (M2.8)	Defensive model (M2.13)
Local increasing returns to scale	$\Lambda = \sum_{j=1}^{n} \lambda_j^* < 1$ for all optimal solutions	$\omega^{\min} \leq \omega^{\max} < 0$
Local decreasing returns to scale	$\Lambda = \sum_{j=1}^{n} \lambda_j^* > 1$ for all optimal solutions	$0 < \omega^{\min} \leq \omega^{\max}$
Constant returns to scale	$\Lambda = \sum_{j=1}^{n} \lambda_j^* = 1$ for all optimal solutions	—

The criteria for characterising economies of scale in DEA as listed in Table 2.1 are based on the solutions of the offensive DEA model under constant returns to scale and/or the defensive DEA model under variable returns to scale (output expansion). The scale factor Λ has already been mentioned in the discussion of Figure 2.6. Economies of scale can also be characterised using the sign of the ω variable estimated by the solution of the defensive VRS model. For example, a DMU with range of $\omega > 0$ in output expansion efficiency operates under local decreasing returns to scale whilst in an input contraction efficiency the criterion operates with reverse signs and therefore range of $\omega > 0$ characterises local increasing returns to scale. In Figure 2.6 the ω variables denoted as ω^I and ω^O respectively correspond to input contraction and output expansion cases.

The returns to scale investigation is illustrated next for the six DMUs used in our example. The relevant information is provided in Table 2.2 below.

Table 2.2
Returns to scale for the six DMU
(Output expansion - Input contraction)

DMU	Model (M2.8) $\sum_{j=1}^{n} \lambda_j^*$		Model (M2.13) $\omega^{max} \leq \omega \leq \omega^{min}$		Returns to scale	
	Output	Input	Output	Input	Output	Input
U1	0.6	0.4	$(-\infty, -2]$	[0.66, 1]	Incr	Incr
U2	1	1.0	-		Con	Con
U3	2	1.4	[0.42, 0.71]	[-2.5, -0.75]	Decr	Decr
U4	3	1.6	[0.62, 1]	$(-\infty, -1.6]$	Decr	Decr
U5	1.6	0.6	1	0.26	Decr	Incr
U6	1.3	0.2	3	0.50	Decr	Incr

DMU U2 is the only technical efficient DMU and therefore operates a most productive scale size. As technical efficient (CRS) DMUs operate under constant returns to scale possible estimation of ω values has not any scale efficiency relevance. For technically inefficient DMUs, however, the estimation of the range of ω values is essential for characterising the presence of economies of scale. It is interesting to observe from Table 2.2 that for managerially efficient but scale inefficient DMUs the ω variable takes multiple optimal values.

4. Managerial implications of DEA

Frontier analysis as a methodology was originally developed for addressing real life managerial questions. Indeed, the very early application of the method by Charnes *et al.* (1978) and (1981) sought to compare the performance of national educational programmes (program follow through) in the USA. Use of frontier analysis models for assessing performance has many by-products with profound managerial implications, Boussofiane *et al.* (1991). We shall discuss next a set of direct implications of frontier analysis whilst more advanced issues are discussed in the next section.

- **Target setting**

The assessment of efficiency using frontier analysis models is based on the projection of inefficient DMUs on the efficient frontier and therefore each inefficient DMU is given a set of input-output targets that would render it efficient relative to the frontier. This feature of frontier analysis is called target setting and as we shall see in chapters three and four target setting is a very important element of organisational control and planning.

The target setting formulae of DEA models are listed in Table 2.3.

Table 2.3
Target setting using DEA

Constant Returns to scale (M2.8)	Input contraction $X^I_{CRS} = \theta^* X_j - s_i^{-*} \quad \vee \quad Y^I_{CRS} = Y_j + s_r^{+*}$ Output expansion $X^O_{CRS} = X_j - s_i^{-*} \quad \vee \quad Y^O_{CRS} = z^* Y_j + s_r^{+*}$
Variable Returns to scale (M2.13)	Input contraction $X^I_{VRS} = \rho^{I*} X_j - d_i^{-*} \quad \vee \quad Y^I_{VRS} = Y_j + d_r^{+*}$ Output expansion $X^O_{VRS} = X_j - d_i^{-*} \quad \vee \quad Y^O_{VRS} = \rho^{O*} Y_j + d_r^{+*}$

The assessment of efficient targets for inefficient DMUs should be seen as independent from the efficiency of the corresponding DMUs. For example, in the constant returns to scale case the input and output efficiency are equal. However, the estimated input and output targets yield different points on the efficient frontier.

- **Focus on performance profile**

The operating profile of a DMU is determined by the amount (size) and mix of inputs employed and the mix of outputs delivered. For a given input/output mix DEA can provide useful information, namely *virtual inputs/outputs* on the extent to which inputs and outputs contribute to the efficiency of assessed DMUs.

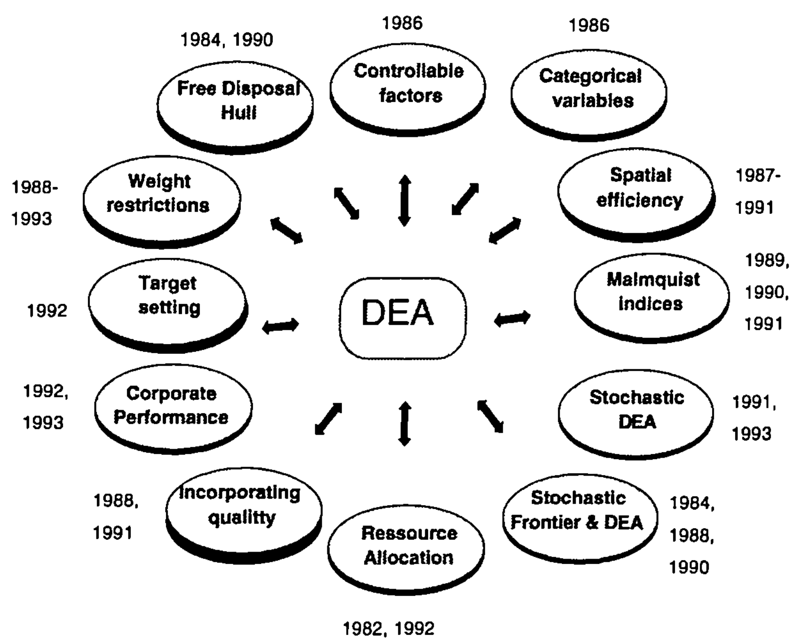
The output expansion efficiency of a DMU is $E_{j_o}^{CRS} = 100 \Big/ \sum_{i=1}^{m} v_i^* x_{ij_o}$, given that the weighted sum of outputs has been standardised as $\sum_{r=1}^{s} u_r y_{rj_o}^* = 100$. The contribution of the i^th^ input and r^th^ output to the assessed efficiency can be defined as $v_i x_{ij_o}^*$ and $u_r y_{rj_o}^*$

respectively. The higher[9] the contribution to the assessed efficiency the better the performance of the DMU is on that particular input/output. A cautionary note must be made concerning DMUs with multiple optimal sets of weights. These are typically relatively efficient DMUs and methods for focusing on their performance are discussed in chapter six.

5. Recent developments of frontier analysis

Frontier analysis, as an econometric and operational research method, has witnessed considerable expansion during the last twenty years. Seiford (1990), provides a comprehensive listing of most of the published and unpublished frontier analysis literature. The number of 500 papers listed in Seiford's literature review indicates the widespread theoretical and applied expansion of the field. The latter, however, makes it difficult to keep up with this expansion in a literature review chapter. A judgmental selection was made therefore to report advanced research issues of frontier analysis since its inception by Farrell (1957).

The chronological tree of frontier analysis



9 This effectively can create problems as units with, say, very low input values would put the majority of their weight on this particular input. In the subsequent discussion we report methods used in the literature for preventing these extreme phenomena.

The review is organised by a thematic area of development without following a chronological time progress.

- **The structure of the Production Possibility set**

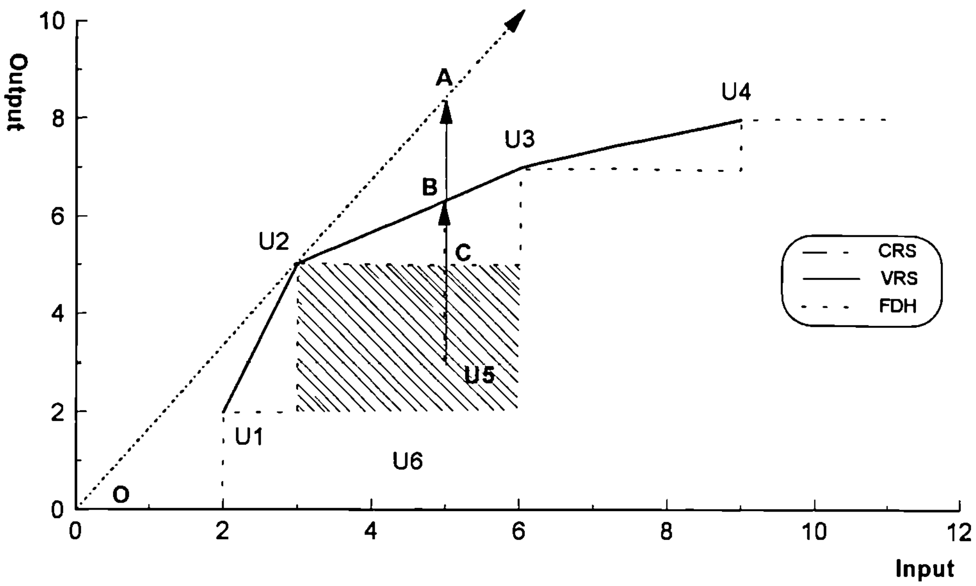
This is perhaps the area where most research effort has been concentrated since the original development of DEA by Charnes *et al.* (1978). The first extension by Banker *et al.* (1984) relaxed the constant returns to scale production possibility set to one of variable returns to scale. Banker and Morey (1986a) move further by making distinctions between controllable and uncontrollable inputs and outputs in assessing technical efficiency. These models were complemented later by Thanassoulis and Dyson (1992). Banker and Morey (1986b) also introduced categorical variables in assigning priorities in the comparisons between DMUs that satisfy given properties. For example, in assessing the performance of restaurants one may restrict efficient restaurants that have drive-in facilities to be compared with restaurants without these facilities but not the other way round. This idea, was generalised later by Dyson *et al.* (1993) introducing the notion of *multiple production functions* within a given production possibility set.

The convexity of a production possibility set was relaxed as early as (1983) by Deprins *et al.* in assessing the performance of post offices in Belgium. This extension received wide publicity and was named Free Disposal Hull (FDH) about seven years later with Tulkens (1990) and his associates at the Centre of Operations Research and Econometrics leading this research direction. The FDH idea is based on the observation that the production possibility set should be made by firms using inputs and outputs without however recognising linear combinations of the observed firms as members of the production possibility set.

The FDH efficient frontier is illustrated graphically in Figure 2.7 using our earlier numerical example in Figure 2.5.

Figure 2.7

Free disposal efficient frontiers



Under the FDH assumption the efficient frontier is made of DMUs U1, U2, U3, U4 which has a step-wise shape. The output expansion required for DMU U5 is now U5C as compared to U5B under VRS and U5A under CRS assumptions respectively. DMU U5 is inefficient as compared with only DMU U2. This is because U5 is located within the area of "dominance" of DMU U2 (shaded area) which assumes that the efficient DMU U2 will always deliver its current outputs if it is provided with more input (disposability). The FDH idea was extended further by Athanassopoulos and Storbeck (1992) in assessing spatial efficiency whilst similar principles have been developed independently by Petersen (1990).

- **Defining efficiency metrics**

Earlier developments of frontier analysis reveal that the definition of production possibility sets and their efficient frontiers should be kept separate from the measure of efficiency of DMUs. To do that one needs to employ "metric functions" that measure the distance between inefficient DMUs and the efficient frontier. A variety of metric measures followed the initial developments by Debreu (1951) and Farrell (1957). One needs to mention the work by Fare and Lovell (1978) which developed the so-called *Russell* efficiency index; the Charnes *et al.* (1985c) attempt which developed the additive DEA model; the Zieschang (1984) attempt to link the Russell-efficiency index with DEA and the Fare *et al.* (1985) and (1993) attempt to define hyperbolic efficiency metrics.

In summary, the main debate on efficiency measures focuses on whether they have a *radial* or *non-radial* nature. Economists like Russell (1985) express the view that efficiency indices should be homogenous of degree-1 and as non-radial measures fail to satisfy this test they are "undesirable". In the management science field Thanassoulis and Dyson (1992) have found advantages in using non-radial efficiency measures for target setting. The advantages of the non-radial efficiency metrics will be capitalised and further developed in chapters four and five of this thesis.

- **Weight restrictions and value frontiers**

The original development of DEA by Charnes *et al.* (1978) was based on the assumption that each assessed DMU should have free choice in selecting weights for inputs and outputs without any preliminary restrictions (see defensive DEA model in M2.11). However, this era lasted until 1986 when Thomson *et al.* (1986) argued that in selecting potential sites for the location of a nuclear research laboratory, they had to restrict the flexibility of weights in order to reduce the number of DMUs assessed as efficient by the standard DEA. This attempt was followed by a very rapid expansion of ideas on how weight constraints should be imposed in assessing efficiency, Dyson and Thanassoulis (1988), Charnes *et al.* (1990) and Beasley (1988), Thanassoulis *et al.* (1994), Dyson *et al.* (1993) and Cook *et al.* (1990).

Dyson *et al.* (1993) in a later attempt sought to investigate the consequences of the use of weight restrictions on the production possibility set of DMUs, the efficient frontier and finally the efficiency metrics. Weight restrictions have so far been used either for reducing the number of efficient DMUs in ordinary DEA or for incorporating experts' opinion of the importance of some of the inputs/outputs on the assessed efficiency. These uses are, however, subject to intense research focus seeking to develop systematic methods for setting weight restrictions and understanding the full impact of weight restrictions on the efficiency process.

- **Integrating the time element into efficiency assessments**

The original use of frontier analysis was based on cross section observations and therefore the efficiency of DMUs was assessed for a particular time period. Charnes *et al.* (1985) introduce the notion of *window-analysis* for assessing performance over time. In window analysis, DMUs with data over a number of time periods are assessed by considering observations from adjacent[10] time periods on one cluster. Next, another cluster of

[10] The number of time periods combined was decided arbitrarily

observations is created by removing the latest time period from the previous cluster and adding observations from one further time period.

The major integration of time into the assessment of efficiency was later developed by Fare *et al.* (1990) which introduced the so-called *malmquist* indices. In the Malmquist analysis a combination of four performance indices are estimated for each DMU. For example if the DMU is observed in time period t and t+1 then we estimate the efficiency of the observed DMU at t against the frontiers at t and t+1; we also estimate the efficiency of the observed DMU at t+1 against the frontiers at t and t+1. Fare *et al.* (1990) and (1992) use the geometrical mean of these indices to define the technological progress/regress of DMUs.

Malmquist analysis has for the last four years been an area of rapid research expansion, Forsund (1992), Lovell (1993). This research has sought to advance the earlier Malmquist models and to decompose efficiency over time into economic components. Attempts to use Malmquist types of analysis for decision support can be found in Athanassopoulos and Thanassoulis (1994) and are discussed in more detail in chapter nine of this thesis.

- **Organisational science, frontier analysis & decision support**

Lewin and Minton (1986) launched an attempt to open a research debate for opening a communication network between the general managerial concepts of performance and organisational effectiveness and frontier analysis. Epstein and Henderson (1988) moved further and examined the appropriateness of frontier analysis as tools for *control* and diagnosis.

At the operational level, Lewin *et al.* (1993) used frontier analysis to support the identification of strategic groups in the brewing industry in the USA. Athanassopoulos and Ballantine (1993) on the other hand worked at the corporate firm level to examine the performance differences between strategic groups in the UK grocery industry, using frontier analysis as the performance yardstick. The early sign of these studies are that frontier analysis can be used for advancing the current state of the art methodologies used in strategic planning. Further research work in this area can be found in the doctoral theses of Gerberman (1992) and Lerne (1992) which are concerned with the use of frontier analysis for addressing strategic management issues.

- **The econometric school of frontier analysis**

The early development of non-parametric frontier analysis in (1978) was followed by a series of studies comparing DEA with traditional econometric techniques. Banker *et al.* (1984), used a known production function with simulated observations to compare econometric with envelopment frontier analysis. Banker *et al.* (1986) also compared DEA with the translog production function in assessing the cost efficiency of hospitals. Ferrier and Lovell (1990) employed econometric and linear programming frontier methods for assessing the cost efficiency in banking. More recently Thanassoulis (1993), has compared ordinary least squares regression with DEA as tools for performance measurement and target setting. These studies have sought to emphasise the differences between econometrics and envelopment frontier analysis and have sometimes resulted in considerable methodological debates between the students of the two methods, (see the debate between Charnes, Cooper and Sueyoshi (1988) and Evans and Heckman (1988) about the break-up of Bell telecommunications due to the violation of monopolistic regulations).

A second and perhaps more interesting field, concerns the attempts for linking econometric and data envelopment frontier estimation in assessing productive efficiency. Here we have two main schools of development.

One of these schools considers DEA as a first stage process where DMUs are assessed for their performance and then are adjusted to their efficient levels (targets) The second stage makes use of econometric models for estimating production/cost functions based on the adjusted (inefficiency free) observations. These type of studies have been advocated by Ray (1988), Tulkens et al. (1993), Cooper and Galliegos (1991) and Sexton *et al.* (1991).

The second school seeks to relax the deterministic nature of DEA in at least two ways. Land *et al.* (1991) and (1993) sought to estimate efficiency by introducing uncertainty into the coefficients of the DEA assessment. This study was based on the chance-constraint approaches developed long ago by Charnes and Cooper (1959). Stochastic DEA has also been developed by Petersen and Olesen (1992) in an attempt to introduce quality dimension in the assessment of performance.

- **Computational aspects of frontier analysis**

The non parametric frontier analysis methods that were discussed in this chapter have a linear programming nature. Thus the computational problems that emerge from frontier analysis can be addressed using the powerful linear programming codes that are commercially available (e.g. GAMS, MPL, AIMMS, SAS/OR, LP88). The solution of separate linear programmes for each assessed DMU requires, however, the generation of a sequence of similar but not identical problems to be solved. Therefore, the computational problems related with frontier analysis are related mainly with speed and efficiency.

During the writing of this thesis there were two commercial software codes specifically designed for non-parametric frontier analysis. The first, is called IDEAS and was developed by Dr. Ali (1989) whilst the second in called the Warwick-DEA (1987) and was developed by the DEA research team at the University of Warwick. A small numerical illustration of the computational efficiency of the Warwick -DEA is given in Table 2.4.

Table 2.4
Efficient algorithmic procedures using the Warwick-DEA
(3 Inputs - 3 Outputs in 486DX processor)

No. of DMUs	Efficient DMUs 10 %	Efficient DMUs 20 %
50	4 sec	9 sec
200	10 sec	24 sec
600	125 sec	220 sec
2000	600 sec	1020 sec

The special structure of a DEA formulation enables the frontier analysis program to develop preliminary tests that identify the dominated DMUs using simple ratio analysis. Dominated DMUs are removed from the basis of the corresponding linear program and therefore the size of the linear programming matrix that needs to be inverted is reduced substantially.

6. Conclusion

The assessment of performance of economic systems has undoubtedly gained substantial publicity over the last two decades. Traditional ways of management and decision making are constantly being revised, and decision support tools are needed to keep up with these changes. Traditional mechanisms of control and performance diagnosis have an accounting bias which do not provide all necessary information concerning the assessment of performance.

Frontier analysis seeks to co-ordinate and integrate research efforts from different disciplines by assessing the productive efficiency of DMUs. The rapid expansion of the field during the last twenty years indicates the healthy prospects of the method in assessing performance. The frontiers of frontier analysis are constantly expanded with applications in previously "virgin" research areas.

The remaining chapters of the thesis will seek to use/extend the existing technology of frontier analysis in assessing the performance of profit making outlets, whilst a new frontier, analysis framework will be developed for linking target setting with resource allocation in non-profit multi unit organisation.

- END OF CHAPTER TWO -

Chapter 3

Target setting as an aid for control & planning in "need based" MUOs

1. Introduction-Motivation

This is the first chapter that deals with performance and planning issues in multi-unit organisations (MUOs) engaged in *a-posteriori* decision making. The flexibility of individual DMUs to allocate their resources with some degree of autonomy presupposes the use of *control* and *co-ordination* mechanisms from central management. The type of organisations discussed in this section, namely MUOs, involve at least two levels of hierarchy. This implies that the analysis anticipates at least two groups of stakeholders, notably central management and individual DMUs' management. For example, central management could represent the governmental policy whilst DMUs could be local authorities. Alternatively central management could be a local education authority (LEA) and DMUs could be the individual schools within the LEA.

In the introductory chapter of this thesis the importance of control and planning mechanisms in MUOs was highlighted. Anthony (1965) defines *control* as the "mechanism that ensures that plans are achieved or goals are attained". To pursue this task organisations employ performance measurement mechanisms. These mechanisms have a reactive role based on the assessment and reporting of past performance. Managerial control, however, can be utilised further if linked with the planning process in MUOs. Target setting is being advocated as the process that could enhance managerial control with planning features. This undoubtedly requires systematic development of the target setting process within MUOs, and will be the main theme of this chapter.

The rest of the chapter is organised as follows. The essential components of performance measurement are discussed whilst particular emphasis is given to the importance of assessing organisational efficiency. The various aspects of goal setting in organisations are introduced and a comparison is made between target setting and efficiency assessments as tools for monitoring performance. The superiority of target setting over simple efficiency measurement is advocated and a set of principles that would enhance target setting as a tool for managerial control and planning are provided. The chapter concludes with an analysis of the operational implications of the target setting principles on the frontier analysis methods introduced in chapter 2. This is expected to highlight the potential extensions to frontier analysis necessary for addressing the new features of target setting.

2. Performance issues in non-profit MUOs

Performance review has always been an essential element in the management of public sector organisations. It is becoming critical today with the increased demand for "improving" the services provided by the public sector. Despite the "popularity" of the notion of performance in public sector activities, little work has been done todate defining the term in its full dimension. The performance measurement manifesto addressed by the Audit Commission (1985) emphasised that performance cannot be captured sufficiently as a single dimension issue. The Audit Commission, and authors in public economics e.g. Brown and Jackson (1990), identify performance as comprising of three successive components, namely Effectiveness, Economy and Efficiency.

Effectiveness measures to what extent an organisation has achieved its targets; to measure effectiveness a set of predetermined ideal targets (outcomes) are contrasted with the actual outcomes of the organisation. Economy measures the cost of achieving the organisational targets; a set of predetermined ideal costs for the resources used is contrasted with the actual costs incurred in measuring economy. Efficiency measures how well an organisation converts its resources (inputs) into outputs and services. This measure helps to estimate the "waste" in organisational operations. To measure efficiency one needs to compare the actual amount of inputs used and outputs delivered between groups of organisations performing similar tasks. A fourth component, namely *equity*, will be introduced later in the thesis as the factor that can be used to link performance (control) with resource management (planning) in MUOs.

The clarity by which performance can be defined theoretically is, however, limited when performance is assessed in practice. The monopolistic nature of many public sector operations does not allow for the development of a *market* which will force prices (costs) to an efficient level. As a result the measurement of economy is a difficult task for management. The definition of *ideal targets* in assessing organisational effectiveness presents similar difficulties which are worsen by the qualitative nature of many public services e.g. health and education. Efficiency appears to be the most plausible performance component to assess as it is based on observed input/output relations. Petisteau and Tulkens (1990), argue that the assessment of technical efficiency (as defined in chapter 2) in the provision of public services is the only defensible though partial indicator of performance comparison in the absence of market prices that would allow for assessing allocative efficiency.

Most empirical and theoretical studies concerned with the quantitative assessment of performance in the public sector seek to assess technical efficiency. This, however, tends to create an imbalance in the importance of other performance components like effectiveness.

The measurement of efficiency in the public sector is by no means a new concept. However, the systematic measurement and reporting of performance, together with its linkage to the decision making process in public sector organisations has emerged during the last two decades in the public management agenda of many industrialised countries. In the UK, for example, performance reviews are more commonly referred to as the *value for money audit* (VFM). The objective of VFM studies is to investigate and improve performance in the provision of public services based on the assessment of effectiveness, economy and efficiency. The instruments customarily used for assessing and reporting performance are the so-called *performance indicators* (PIs).

Performance indicators are univariate measures that quantify the relationship between inputs and outputs of operating units. At present they cover the majority of public sector activities in the UK and some thousand PIs are currently used to assess performance of public sector activities. The introduction of PIs in the public sector is used to satisfy two main objectives. The first is internal and concerns the management control process in the provision of goods and services. It is anticipated that constant provision of information on the performance status of organisations will increase the management's awareness for the need for better efficiency. The second objective is external and concerns the use of PIs for informing taxpayers on the level of efficiency in the provision of public services. Here the use of performance measures anticipates the "benefits" from external pressure that can be imposed by tax-payers demanding higher performance of public services.

Examples of external use of performance assessments can be found in the publication of comparative statistics concerning the provision of local services by local authorities in the UK and also the publication of the so-called "league tables" concerning the performance of schools in the UK. Butterwoth *et al.* (1989) and Smith (1993) argue that the comparative statistics of that kind have not been welcomed by the interested parties either because of methodological problems or because they fail to provide information of the vital parts of the activities of the organisations assessed (e.g. information on outcomes).

There is a host literature which criticises the use of PI as a general concept (Smith (1990), Mayston (1985)) or for specific sectors of public activity (Boussofiane *et al.* (1990), Birch and Maynard (1986)). Furthermore, Smith (1993) highlights the importance of outcome related performance indicators and discusses intended and unintended managerial and behavioural consequences from the introduction of performance measurement in public services. This work was extended later by Dyson *et al* (1994) which provide a managerial framework under which DEA can overcome the limitations of tools like the PIs.

The critique of performance indicators concerns their nature and also their managerial use. PIs have univariate nature and therefore concentrate on one performance dimension at a time without providing a global performance perspective on the assessed performance of units. In a situation where, say, five PIs are used to assess performance of a set of DMUs, each DMU will be assessed a target value for each PI. These targets are not realistic as they reflect the extreme performance of different DMUs on each of the five PIs separately.

The limited ability of PIs to capture overall performance makes them appropriate for only diagnostic studies without any scope for supporting decision making. The use of PIs for assessing performance in the public sector may have partial success to identify some *symptoms* rather than the *causes* of poor performance. Although the identification of the real *causes* of under performance is a difficult problem, other methods e.g. regression and data envelopment analysis are advocated by Sherman (1986) and Smith (1990) as being more insightful than performance indicators. In brief, this is mainly because regression and data envelopment analysis can accommodate multiple inputs and outputs; provide summary information on the overall performance of assessed DMUs and finally investigate important issues concerning returns to scale and input/output rates of substitution, Thanassoulis (1993).

The theoretical debate on the selection of the most appropriate performance measurement tools has recently been enhanced by the governmental tendency in the UK to utilise performance for decision support[1]. This new role of performance measurement, however, reopened the debate on the relation between efficiency the other performance dimensions, namely effectiveness, economy and more importantly equity as resource planning criteria.

Clearly equity in the provision of public goods is a very important issue that has been addressed mainly from the political science, Lineberry (1977) and economic literature Thurow (1976). The distribution of public services can be formulated using alternative equitable formulae with different policy making nature. Savas (1981) proposed a classification of equity formulae into four basic categories: equal payments, equal outputs, equal inputs, and equal satisfaction of demand. These principles represent alternative considerations on the criteria that an allocation process should follow in order to distribute resources equitably. Unfortunately, there is no unique way to define and measure this process. One may argue that public services should be allocated in proportion to what people are prepared to pay for services they receive from providers. In the case of road lighting or sanitation this may be justifiable but when a house is on fire many people other than the homeowner benefit from having the blaze extinguished before it spreads. Many examples and counter examples can by found in Savas (1981) highlighting the importance and complexity of defining equity as a criterion for resource management in the public sector.

The concept of equity and its relation with efficiency and effectiveness will be explored in more detail in chapter six of the thesis. The intention in this chapter is to highlight the tradeoffs between efficiency and equity when they are considered as objectives for allocating resources. Whilst in the seventies these trade-offs were mainly discussed within the academic community the political changes in the UK during the eighties have brought these tradeoffs to the forefront within the managerial agenda of real life decision making. At the extreme one can find two opposing views on efficiency and equity.

---

1 Attempts are on their way in the UK to link performance and resource allocation in series on public activities e.g. Local authorities, Schools, Universities, Police forces.

Those that advocate *efficiency* as the most important criterion in the provision of public services often adopt cost saving policies. They argue that the lack of clear efficiency objectives results in high resource underutilisation which in turn puts under pressure the fiscal policy of central governments. On the other hand, the advocates of *equity* and relative need are mainly concerned with reduction of inequality in the allocation of resources. There is an ongoing concern, for example, on the impacts of the internal markets in the national health system in the UK on the equity by which health services are provided.

This thesis will acknowledge the importance of both issues as vital components in the planning process within MUOs. Efficiency, effectiveness and equity will be recognised as objectives of MUOs' management and therefore any resource planning process would seek their satisfaction. The use of target setting is intended to facilitate the quantitative representation of these objectives in the planning process whilst to support anticipating the tradeoffs concerning their satisfaction. This would facilitate to overcome the polarisation of the objectives of resource management within an integrated planning model.

3. Assessing performance through target (goal) setting

As mentioned earlier, target setting is adopted in this thesis as an appropriate tool for managerial control and planning in MUOs. An introductory discussion on the nature and classification of target (goal) setting in organisations is made prior to investigating its usefulness as a performance measurement tool. As Brown and Pyers (1988) argue target and goal setting constitute integral parts of performance measurement, and in some cases can motivate organisations for improved performance. It is also emphasised, that target setting in the public sector is underutilised and, moreover, it is analogous to the profit motive found in the private sector.

Organisations are goal-attainment devices. An organisational goal is a desired state of affairs that the organisation attempts to reach, Amitai Etzioni (1964). A goal represents a result or an end point toward which organisational efforts are directed. There are two broad categories within which organisational goals can be classified: the officially stated goals of the organisation; and the operative goals that the organisation actually pursues, Daft (1989).

Table 3.1
Classification of goals

Type of Goals	Purpose of Goals
Official goals Mission statement, Vision	Legitimacy, Public image
Operative goals Overall performance, Productivity, Increase quality of services	Employee direction & motivation, Monitoring decision making, "Accountability" over time, Reduction of uncertainty, Performance standards, Integrating control and planning

Table 3.1 exhibits the two goal categories together with their purpose. Official goals are often found to be abstract and vague. They cannot be measured precisely, and goal attainment cannot be evaluated precisely. This is because official goals describe a *value system* for an organisation; they represent broad vision, and they seek to legitimise the organisation. The "mission statement" usually embraces the various stakeholders in the organisation and highlights quality, responsiveness, competitiveness, job satisfaction, community well-being and ecological concerns.

The use and publicity of the official set of goals is typically found in private sector organisations as a result of market competition and the pressure to keep up with shareholders' expectations. The statement of official goals in public sector institutions has gained publicity in recent years as a result of the increased demand for public accountability. Examples of official goals in the public sector can be found in the citizen's charter, the white paper for health care provision in the UK, the Maastrict treaty for the European Union, the election manifesto of each political party before the election, etc.

Operative goals designate the ends sought through the actual operating procedures of the organisation and explain what the organisation is actually trying to do, Perrow (1961). Operative goals describe specific measurable outcomes and they pertain to the primary tasks an organisation must perform. Table 3.1 lists some examples of operative goals. The quantification of the overall performance of organisations can be expressed in many different ways. Pittsburgh Plate Glass company, for example, had defined as an operative goal raising return on equity to an average of 18% in 1994 compared with 15.7% in 1984, and increasing annual sales volume to $ 10 billion from $ 4.7 billion, Resener (1987). The reduction of the public deficit as a percentage of the GDP in the UK in the period 1992-1997 is an operative goal in the public sector. The creation of some thousands of new jobs in the European Union via infrastructure investment is an operating goal for combating unemployment.

Service level goals mirror the public image of not-for-profit organisations. For example, response rate by emergency units, delivery rates by post offices, punctuality of trains, time spent on the NHS waiting lists, etc. constitute service goals for individual public sector organisations. Achievements in the level of provided services are often used in the public debate concerning the performance of public sector organisations. *Technology* goals relate to organisational targets to adapt to market changes by launching new products, delivering new services and developing new production processes. R&D expenditure is considered as a primary criterion affecting the long run viability of organisations and, therefore, it is reported annually for all sectors of the economy in the UK, (the UK R&D Scoreboard). Finally, *productivity* goals are concerned with the amount of output achieved from available resources. In periods of recession productivity goals tend to be key factors of organisational viability.

Official goals have a very broad character and they reflect organisational vision at a strategic level. Therefore, official goals make no specific quantitative references as to *when*, *to what extent*, and *how* they will pursue their targets. Operating goals cover all aspects of organisational activities in a more specific and quantified manner.

> ***The development of models for estimating targets for <u>operating goals</u>, as well as facilitating their achievement is the primary aim of this research.***

Having provided an introduction to the notion and classification of organisational goals we can now turn to discussing the association of goals and targets with organisational performance.

3.1. Efficiency assessments and target setting

Field and Shutler (1991) investigating the use of performance targets in eight nationalised industries in the UK argued that there is plenty of detail concerning the actual performance but little concerning targets. Along the same lines Allen (1994) recognises the inherent problems in attaching quantitative values in setting targets. Field and Shutler also argue that it is difficult to isolate the various performance components, namely economy, effectiveness and efficiency, from the annual reports of private/public organisations. This may limit the ability to compare expected with actual performance to only financial factors.

Companies that employ target setting (e.g. Post Office Counter Services) do not necessarily adopt systematic procedures for monitoring and reviewing this process. This would assist organisations into setting feasible and challenging targets without over or under estimating the ability of the organisation to meet these targets. Moreover, action scenarios need to be put forward supporting the achievement of the assessed targets. For example, the targets concerned with the volume of letters processed and/or customers serviced by the Post offices can be affected by the extent to which investment is made for training, automation and new technology.

The association between performance measurement and target setting is unclear. Despite their affiliation they tend to be parts of unrelated processes, namely control and planning and therefore used for addressing different organisational problems. The literature discussing performance issues in the public sector, e.g. Smith and Mayston (1987), Smith (1991) focus on the pros and cons of tools for assessing performance. Methods such as *performance indicators*, *regression analysis, data envelopment analysis* have been proposed, applied and criticised for assessing organisational performance.

As mentioned in chapter two the more advanced of these methods, namely data envelopment analysis, yield as by-products "efficient performance targets". However, these performance targets are treated more as by-products rather than as the main objective of these studies. The inability of performance indicators to be used for setting targets has been noted earlier in this chapter. On the other hand, the claim that the current frontier analysis literature gives secondary importance to the estimation of targets is a fair one. Exceptions to this trend are the frontier analysis studies conducted by Golany (1988), Thanassoulis and Dyson (1992), Thanassoulis (1993), Thanassoulis *et al.* (1993), Athanassopoulos (1994a) and (1994b) that have a "target setting" purpose.

Some closer investigation is needed for looking at the relationship between efficiency and target setting. Target setting has fundamental differences compared to the efficiency assessment methodology as can be seen in Table 3.2. Efficiency measurement studies are considered as "external audit" of various organisational functions. Needless to say, the assessed management regard this process as a potential "threat" and therefore, is more likely to resist instead of collaborating in a performance measurement exercise. On the contrary, had management been consulted in formulating targets based on mutually accepted objectives, performance measurement would be integral to both the control and planning processes.

Table 3.2
Target setting Vs. efficiency assessment

Target Setting	Efficiency Assessment
Comprises Control and Planning dimensions	Has a predominate control dimension
Closer to organisational objectives	Perceived as an external process from management
There is internal and external "accountability" over time	Major role is to discriminate between "good" and "bad" past performance
Requires close managerial consultation	Difficulties to convince management of the practical value of efficiency assessments

An efficiency assessment study seeks to discriminate between good and bad performers at a particular point in time. However, this evaluation of organisational units does not provide any insights into the nature of the assessed inefficiencies nor does it recommend strategies for improving efficiency.

It can be argued that target setting is closer to the organisational strategy/objectives as it is a natural way for their quantitative representation. The formation of targets requires management participation which in turn enhances the validity of the target setting outcomes as compared with those from the efficiency assessment. Target setting also provides an accountability framework where organisational efforts for achieving the estimated targets can be examined over time. Accountability related to efficiency improvements in organisations, on the other hand, gives very limited information on whether individual units' performance changes over time are due to their efforts or due to the variation of the standards they get compared with. This is because the standards are based on the best observed practices of each time period which do not remain constant. The advantages of target setting will be exhibited in more detail in the discussion of the principles of effective target setting.

3.2. Principles of effective target setting

The discussion above on the differences and similarities between target setting and efficiency assessments suggested that target setting is a more advantageous approach for improving performance. However, the discussion also emphasised the lack of systematic processes for setting targets that would encapsulate control and planning dimensions. This can be facilitated by developing a set of target setting principles which state the conditions that need to be satisfied if targets are to be useful in *control* and *planning* processes.

Table 3.3

Principles of Effective Target Setting

1.	Operational *feasibility* of the assessed targets
2.	*Value revelation* principle (Multiple objective dimension)
3.	Incorporating managerial preferences over input/output improvements
4.	Reflect desired global organisational performance
5.	Develop supporting mechanisms to monitor the target's achievement

The set of "principles" in Table 3.3 is intended to open a very important debate regarding the enhanced role of target setting as a *control* and *planning* instrument. The "debate" has two basic dimensions; firstly it explores the characteristics and appropriateness of each of these principles, and secondly it develops operational models for estimating targets that satisfy these principles.

3.2.1. Operational feasibility of the assessed targets

The term "***feasibility***" in the context of target setting has a twofold interpretation. The first emanates from the actual comparative basis that targets are obtained, whilst the second relates to the levels of estimated targets.

Smith (1990) criticises the use of PIs arguing that they do not provide a systematic basis of comparator practices that would allow target estimation. The characteristics of the comparator groups is a topic with limited concern in the performance measurement literature. The selection of appropriate comparators, however, can increase management's confidence on the assessed targets and also increase the likelihood for achieving these targets.

Use of the *frontier* based methods (DEA), as introduced in chapter two, yield as a by-product targets for underperforming units. These targets are made of combinations of "best practice units" (efficient). The selection of "best" units determines the nature and magnitude of these targets. Management is given the option to validate the assessed targets by inferring whether the units used as comparators have compatible operating profile to the targeted unit. This validation process is based on the notion of *perceived fairness* concerning the way management sees any performance measurement process. Epstein and Henderson (1989) rated perceived fairness as a critical factor that affects the success of any performance measurement system. In the frontier analysis literature steps for strengthening the fairness in the assessment of performance can be found in the models formulated by, Tulkens *et al.* (1992) and Athanassopoulos and Storbeck (1992).

The level of assessed targets needs to be *realistic* as well as *challenging*. Operating units given targets to deliver the current levels of services at, say, 60% lower costs cannot usually be expected to achieve these targets unless some supportive actions take place. The case of unrealistic performance targets is seen in line with the "nature" of the observed inefficiencies of operating units. It is noticeable that the *nature* of the observed inefficiencies in the current literature of performance measurement attracts technical attention such as "technical", "scale", "allocative", "scope", and "program" efficiencies, see Charnes et al. (1981), Banker et al. (1984) and Fare et al. (1985). It is important to mention again Leibenstein's (1976) contributions introducing the concept of *X-efficiency* in order to determine the causes of under performance of firms/units. Some light is expected to be cast on this issue in the discussion of the *value revelation principle* which now follows.

3.2.2. *Value revelation*

Target setting was acknowledged earlier as being closer to organisational objectives in comparison to efficiency assessment studies and therefore it can guide the assessment of organisational effectiveness. The assessment of targets therefore should encapsulate the organisational mission and the subsequent objectives used for its quantification. This would aid the *revelation of organisational values* as far as the performance of the organisation is concerned.

To accommodate organisational objectives in setting targets one needs to find quantifiable factors that will describe the operations of individual units. An investigation can be made assessing the potential improvements of these factors in line with the managerial/economic objectives (e.g. input reduction) of the organisation.

The principle of *value revelation* has similarities with the notion of *triangulation* of measures and methods. The concept of triangulation is simple: **Any measure in isolation can be misleading**, Welch and Comer (1988).

> *Multiple measures that use variations in conceptual perspective reduce the possibility for invalidity and provide a more comprehensive assessment of the concept.*

Departures from the aforementioned principle result in selective choice of measures that are optimistic or pessimistic assessments of performance depending on the purpose of the performance review exercise.

In the UK the bulk of the current performance indicators used in the public sector have a predominate cost saving orientation, Smith (1990) and (1993). The complexity in the operation of public sector MUOs, however, advocates the development of more representative systems to accommodate organisational objectives in the target setting process.

A target setting process at a school, for example, would need to consider the mission statement of the school (or the Local Educational Authority) and identify criteria (inputs and outputs) for its assessment. This will imply inclusion of curriculum and extra curricular achievements of pupils, the sociodemographic background of pupils and their abilities prior to entering the school. Having taken these factors into account the target setting process would support the identification of the strengths and weaknesses (e.g. value for money and value added) of individual schools as far as their mission statement is concerned.

3.2.3. Incorporating preferences over input/output improvements

The representation of the organisational mission in the inputs-outputs of target setting was established as a target setting principle in the discussion of value revelation. This principle, however, needs to be enhanced by incorporating *managerial directions* within the target setting processes.

The use of managerial preferences can also aid in accommodating strategic planning issues into the assessed targets. Performance measurement systems are dominated by the accomplishment of the present in comparison to the past, while usually there is little or no explicit concern for the future. Estimation of performance targets based on best observed behaviour of previous years assumes that the organisational strategy remains stable over time. **Undoubtedly, poor performance achievements in the past are sometimes attributable to inadequate organisational strategy. Neglecting this possibility may lead to reproducing the causes of poor performance and estimate performance targets non-compatible with a revised organisational strategy.**

In an organisation that uses multiple inputs to deliver multiple outputs there are tradeoffs regarding the *orientation* of the assessed targets. In the assessment of targets for schools, for example, the orientation given to these targets has profound implications on the operation of each school. On the outcome side there are questions on the extent to which the academic achievements should be considered as more important than extra curricular activities (e.g. music and sports) or whether the placement of 15-16 years pupils to work will be considered as a preferred direction to having pupils with low GCSE achievements. Other important issues concerned with the direction of the assessed targets emanate from the tradeoffs between resources and outcomes at schools.

3.2.4. *Reflect desired global organisational performance*

MUOs have a multi-unit, multi-level structure. The target setting process needs to respect the organisational structure of MUOs and thereby assess targets suitable for different levels of decision making. Targets focusing at the global organisational level are of particular importance for an organisation as they represent a direct and quantitative representation of their mission statement which is recognised by Mandell (1991) as a surrogate measure of organisational effectiveness.

The mission of corporate operational goals is closely related with the quantitative representations of the legitimate (official) goals of organisations. For public sector organisations official goals are often determined externally. In the UK, for example, the "citizen's charter" provides a manifesto for the provision of public services. Sector specific *governmental acts* provide more specific policy directives over the goals and mission of organisations such as the National Health System (NHS), Education, etc. The quantitative representation of these official goals are typically of a financial nature and are stated through annual budgets which also have external influence e.g. treasury department.

Global organisational targets cannot always be decomposed on a *pro rata* basis for each operating unit whilst on the other hand, aggregation of targets assessed for individual units is not an appropriate method for their assessment.

An underlying assumption of the global target assessment is the principal-agent relation that links central and individual DMU management. Performance targets in a principal-agent context may result in resource reallocation among operating units. The latter constitutes the principle of *transferability* of resources and performance targets among individual units.

Figure 3.1 shows the framework under which global target setting can be monitored by the assessment of DMU based targets.

Figure 3.1
Global targets and resource transferability

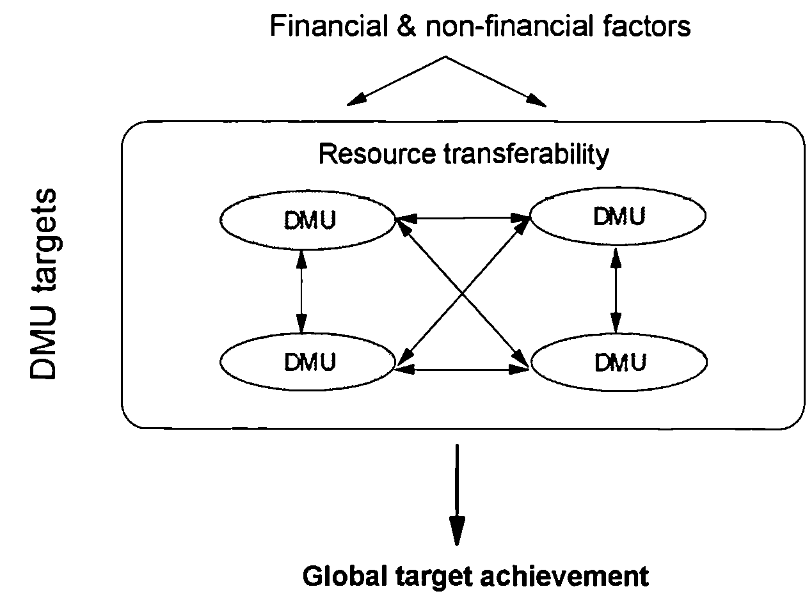


Figure 3.1 gives an essential role to resource transferability as being the instrument that would facilitate the achievement of estimated targets for individual DMUs and furthermore for the whole organisation. It can be argued that other mechanisms can also be employed to foster the targets' achievement (e.g. reorganisation, incentives and training).

3.2.5. Develop support mechanisms to monitor the targets' achievement

MUOs give high priority to the development of effective performance review systems. Information systems are developed over time to support private/public organisations being accountable to their *principals.* The intensive interest of the investors (principals) to examine the degree to which management (agents) utilise their capital has advanced very powerful information systems of financial reporting. Therefore, financial reporting operates as a stimulating mechanism towards the objectives' achievement in the private sector. Moreover, the investor's discretion to take immediate action with regard to the reported performance increases the value and use of performance measurement[2].

---

2 Phenomena of "creative accounting" can detract financial reporting from relevance and reliability.

In the public sector, performance reporting is realised as a type of information flow with no immediate impact on managerial decision making. The recent governmental initiatives in the UK[3], to link performance with resource allocation decisions creates incompatibilities between the reporting purpose of the current performance measures and the requirement of performance based resource allocation.

The interchangeable role between tax payers, government, local governments and the employed management as *principals* and *agents* create extreme complications. It is not feasible, therefore, to have analogies of <u>sources</u> and <u>destinations</u> for the concept of *accountability* in the public sector. This complication may increase due to the absence of monitoring mechanisms for adjusting the performance of individual units.

The assessment of performance targets has not much to contribute to organisations that are not prepared to support the whole target setting process. In the best case the assessed targets will be included in the annual budgeting reports. It is argued here, however, that the incorporation of target setting within the decision making process is a *necessary* condition for achieving the assessed targets. The most effective decision making mechanism in the public sector is the resource allocation process, and therefore, the targets' achievement should be monitored in line with resource allotments in public organisations.

Summarising, it is argued that the aforementioned principles of target setting determine the conditions for using target setting as an effective tool for managerial control and planning. The development of operational tools that will support the assessment of targets compatible with the principles of target setting is one of the main concerns of subsequent chapters.

4. Managerial & operational implications of target setting

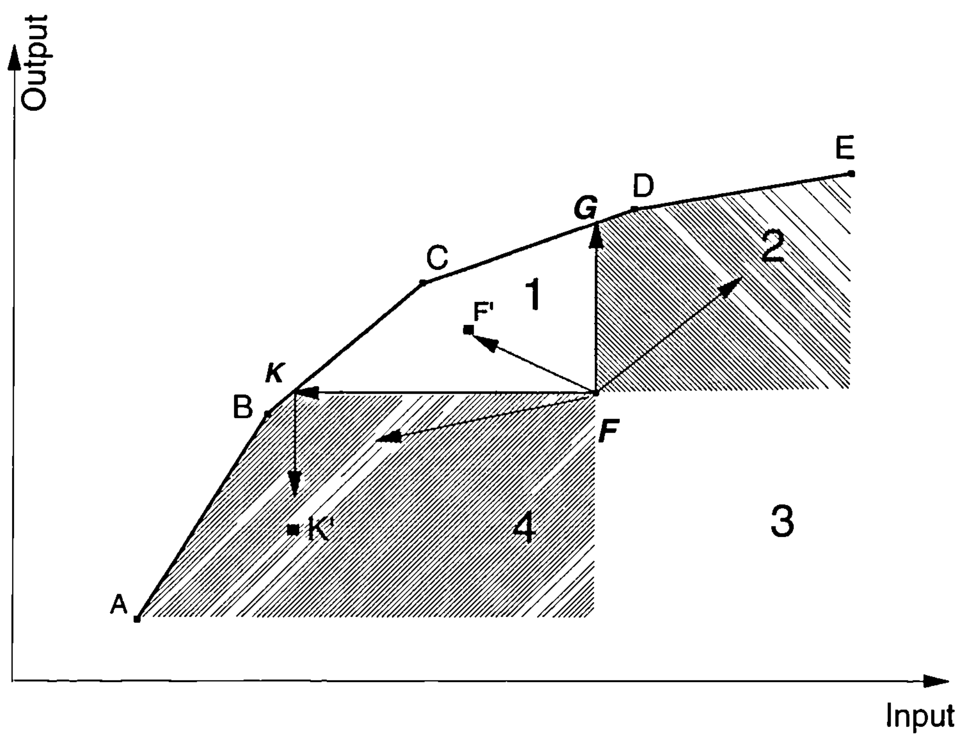
The assessment of performance targets complying with the principles of effective target setting is, undoubtedly, a challenging issue for MUOs. These principles have a variety of implications for the current methodologies of target setting. The most advanced of these methodologies is based on the notion of frontier analysis introduced in the second chapter of the thesis. It is important, therefore, to examine the extent to which these methods accommodate the principles of effective target setting.

---

[3] A clear case of this kind of this kind of initiatives can be found in the allocation of governmental grants to higher education institutions in the UK.

To facilitate our discussion a graphical illustration of a single input/output production technology Figure 3.2 will be used. Units A, B, C, D and E are relatively efficient units (under variable returns to scale) whilst unit F is a relatively inefficient unit located in the interior of the production possibility set. The assessment of performance targets for unit F will be the main theme in our discussion. The traditional DEA philosophy for assessing targets as introduced in chapter two will be enhanced so as to address more demanding features introduced via the target setting principles.

Figure 3.2
Target Setting strategies



4.1. Projections on the efficient frontier

Unit F would be projected to a point on the segment KCG of the efficient frontier using DEA to assess its performance. An output expansion orientation will project F at point G whilst an input contraction at point K. These two strategies seek to improve in one dimension (input or output) by keeping fixed the other dimension. More advanced DEA based targets can be obtained by projecting unit F at alternative points of the segment KG. Thanassoulis and Dyson (1992) developed an extension which yields targets involving input and output improvements simultaneously. The choice of mixed target setting strategies has undoubtedly very important managerial implications as, for example, the alternative points, say point C, along the efficient segment KCG reflect simultaneous improvements to the input and output values.

There are a number of issues related to the simultaneous input/output improvements in assessing performance. First, an element of choice prevails concerning the decision makers' priorities on the improvement of inputs and outputs. On the other hand, assessment of performance with simultaneous consideration of inputs/outputs can be used to address operationally the value revelation principle introduced earlier on. Finally, the piece-wise shape of the efficient segment KG characterises different rates of substitution (slopes) along the segments KC and CG. In brief the rates of substitution give a measure of cost/benefit for substitutions between the inputs and outputs of DMUs that are lying on the efficient frontier. The use of the rates of substitution for setting targets and allocating resources are discussed in more detail in chapters four and six.

The previous target setting strategies yield target values that would render unit F efficient. Let us assume that an input contraction strategy has been adopted for unit F. This implies that the unit is expected to deliver its current level of outputs by using the amount of resources that corresponds to point K. The *mathematical feasibility* of such a projection, however, does not guarantee *managerial feasibility* too. A reduction to the input levels of unit F could perhaps have a negative impact on its performance by forcing it to a point such as to K' in the near future. The target setting strategy, therefore, needs to consider the expected likelihood of achieving the estimated targets based on close managerial consultation. This implies that the target setting process should be accompanied by extensive inspection on the intentional/unintentional consequences of the proposed targets.

The assessment of performance by focusing on individual operating units does not provide any information concerning the global operations of the organisation. In a MUO structure, units share resources, information and goals whilst retaining their individual characteristics. Assessment of performance targets for individual units represent their contribution towards the achievement of global organisational targets. However, the individual DMUs' contribution can be assessed and maximised only if there is an independent process that will proceed to estimate global organisational targets. The assessment of global organisational targets and its linkage with the assessment of DMU based targets will be discussed in more detail in chapters four and five.

4.2. From efficient frontier to improved directions

The transition from a <u>unit specific</u> to an <u>organisation specific</u> target setting process requires a rethink of some of the underlying assumptions of the *efficient frontier* philosophy. Let us consider again the case of unit F. In our previous discussion it was argued that the segment KCG should be considered for obtaining efficient targets for unit F. Two important questions arise from the simultaneous consideration of unit and organisational based target setting.

- Is it possible, in practice, that all inefficient units should be projected onto the efficient frontier ?

Clearly, some units will improve and some will worsen their performance over time as a result of managerial decisions, innovation, as well as uncontrollable environmental changes. The frontier analysis literature, however, is concerned mainly with the assessment of performance targets and not with the follow-up process which will support the achievement of these targets.

- Is there a trade-off between the per unit and the global organisational target achievements ?

The achievement of global goals does not always translate automatically to stretching all operating units to their maximum achievable performance level. The assessment of performance targets for electricity generating plants would not consider all plants producing at the maximum of their capacity. The relatively cheaper units may work overtime whilst the more expensive plants may operate under capacity underutilisation. The criterion for these decisions, however, is the way the central organisation selects its strategy for generating electricity to meet its demand. One also needs to consider the effects of organisational slack as discussed by Bourgeois (1981) which indicates that one needs to take into account the tendency of individual DMUs to create a tolerance level of performance (slack) in their operation.

Unit F, for example, could well improve its performance and operate at the levels of unit F insofar as the global organisational performance is concerned. In general, all points within the shaded area FKG correspond to potentially improved directions for the inefficient unit F. These improved directions yield positions that *dominate* the original position of F but are, nevertheless, sub-optimal solutions in view of the efficient frontier. The possibility of exploring sub-optimal strategies for individual units should be in line with the global organisational strategy of goal achievements.

4.3. Extended Pareto efficient strategies

Hitherto, it was argued that the assessment of DEA targets for individual units can be enhanced by considering *improved direction strategies* which seek to assess global organisational targets alongside the targets of individual units. These policies aim to explore all possible scenarios that improve the units' performance on the basis of individual and/or global organisational criteria. The improvement directions, efficient or not, are considered within the shaded area 1 of Figure 3.1. This is a restrictive strategy which emanates from the way traditional DEA assessments are employed; the comparator efficient set needs to dominate an inefficient unit in all input/output dimensions.

Using this policy for setting performance targets we exclude two areas of the production possibility set that could be considered as feasible directions for setting targets. Using area 2 for estimating targets for unit F would result in targets that seek to augment the unit's output by augmenting its resources. Targets from area 4 on the other hand seek to reduce the scale size of operation for unit F by reducing the input it uses and the output it generates.

The possibility of projecting DMU F to area 2 or 4 introduces a *transferability* and/or *internal communication* feature in assessing performance targets. Units may acquire higher levels of resources in anticipation of higher levels of returns; strategic organisational choices may wish to contract the input/output operation for selected groups of operating units (reduced capacity utilisation). Clearly the encapsulation of communication features between individual units in assessing performance targets is not within the capability of the current DEA models.

The introduction of the principles of effective target setting and the subsequent discussion of the geometrical illustration of Figure 3.1 leads into two main conclusions.

- Traditional target setting methods need to be extended in order to address the extended purpose of target setting as introduced in the principles of target setting.
- There is a linkage between the target setting and the resource management processes followed by MUOs at the DMU and global organisational level.

Chapters four, five and six that follow concentrate on the development of operational models for assessing targets compatible with the target setting principles at the individual unit level. Issues related with the economic and managerial implications of these models will be used in the next chapters for addressing the issues of global target setting and target based planning models.

5. Conclusions

This chapter sought to address target setting in MUOs. Particular emphasis was placed on not-for-profit MUOs where the assessment of performance targets seems to be a more involved issue. Typically, not-for-profit organisations operate in an *a-posteriori* basis where DMUs manage their resources with some degree of relative autonomy. The systematic assessment of performance targets in organisations that give wide discretion to individual DMUs to allocate/manage resources affects the performance of organisations as a whole. Target setting is perceived, therefore, as a mechanism that would aid the managerial *control* and *planning* processes in MUOs.

A set of principles were put forward for target setting in MUOs in order to strengthen its relative importance as control and planning tool. This framework reflects the fundamental characteristics that target setting should accommodate in order to have an effective role in MUOs. The possibility of extending the capability of traditional target setting tools like data envelopment analysis was also discussed. Particular emphasis was given on the incorporation of managerial preferences during the target setting process and also the linkage between DMU based and global organisational targets. This task is pursued operationally in the next chapter.

- END OF CHAPTER THREE -

Chapter 4

Analytic tools for assessing performance targets & their economic implications

1. Introduction

Chapter 4 builds on the issues raised in chapter three regarding the assessment of performance targets in MUOs. In chapter three, it was argued that target setting is currently underutilised as an effective decision support tool in MUOs. Attempts were made to highlight the advantages of adopting target setting procedures within managerial control and planning systems. This idea was developed more systematically by proposing a set of principles for effective target setting whilst the investigation that followed focused on the ability of methods like data envelopment analysis to accommodate these principles.

In this chapter the main objective is to develop operational models which encapsulate the principles of effective target setting as discussed in chapter three. Traditional performance measurement tools in not-for-profit organisations, notably performance indicators, are characterised in the literature (see chapter three) as inappropriate for setting performance targets. Therefore, this chapter will consider frontier analysis as the means for estimating performance targets.

The rest of the chapter is organised as follows. The development of target setting models originates from the individual unit level. Emphasis is given to developing models that encapsulate the principles of target setting highlighted in chapter three. The models developed for assessing targets at the DMU level will be extended to assessing targets for the global organisation. Next, the managerial and economic implications of the proposed target setting models are explored. This leads to a discussion of the economic notion of substitutability and its relation with the target setting models. The economic implications of the target setting models are elaborated further using a small numerical example. The chapter concludes with reference to the ability of the models developed to satisfy the principles of effective target setting that were developed in chapter three of the thesis.

2. Setting performance targets for individual units

The fundamentals of target setting in MUOs were outlined in chapter three. The debate on the assessment of targets continues, however, emphasising the quantitative estimation of DMU and global based targets.

In chapter one it was argued (see section three) that there is a hierarchical link between resource management and target setting at the unit level of MUOs. Resources are allocated to individual DMUs to enable them to achieve performance targets. Target setting strategies need to be developed taking into account the following issues.

- On the <u>input</u> side, not-for-profit MUOs depend on governmental choice of policies. A tight budgetary control policy on a particular fiscal year would automatically give a cost saving orientation to the corresponding DMU based targets.
- On the <u>output</u> side, targets are mainly *official* goals included in the mission statement of organisations. The actual implementation of output targets, however, cannot be based on political intentions and needs to be supported by operational procedures.
- Provision of public services is monitored by legislative rules which enforce their provision without considering marketable benefits. For example, the operation of schools in areas without sufficiently large numbers of pupils can be protected by legislation in order to enforce equity principles.

The issues raised above affect the estimation of performance targets at the individual DMU level and need to be incorporated in the estimation process. The estimated targets will be based on relative performance comparisons in the absence of absolute efficiency measures. Data envelopment analysis (DEA) will be the method used for estimating these targets.

2.1. Current association between DEA and target setting

The use of DEA for setting targets has normally been seen as a complementary stage of the efficiency measurement study. In practice, many analysts have found more effective to communicate DEA results by converting the efficiency measures into target equivalents. It is argued, however, that despite the usefulness of current DEA-based targets for communicating efficiency results to decision makers, they do not sufficiently accommodate the principles of target setting introduced in chapter three.

Given a set of 1,...j..,n operating units we shall denote $X \in \Re_+^m$ a vector of input quantities 1,...i...,m and $Y \in \Re_+^s$ a vector of output quantities 1,...r...,s. Combinations of these input/output vectors generate a production possibility set (see Chapter two, section 3.3) while the *Pareto-undominated* units of this set define its efficient frontier. DEA models seek to identify best practice DMUs and thus derive an empirically based implicit production function for the entire set of operating units. The solution to a DEA problem associates every operating unit with an *efficiency metric*. This metric measures the distance between the operating unit and the efficient frontier of the production set.

In an attempt to link efficiency assessment and performance targets Charnes *et al.* (1978) and Charnes *et al.* (1985d) use the efficiency rating of DMUs to obtain performance targets. The efficiency projection formulae, namely input contraction and output expansion, under an assumption of constant returns to scale are provided in M4.1. The estimation process for these targets has been described in chapter two and is based on the solution of the LP models in M2.8.

$$\begin{array}{lll} \textit{Input contraction} & X_{CRS}^{I} \xrightarrow{frontier} \theta^{*} X_{j} - s^{-*} \textit{and } Y_{CRS}^{I} \xrightarrow{frontier} Y_{j} + s^{+*} & \\ \textit{Output expansion} & X_{CRS}^{O} \xrightarrow{frontier} X_{j} - d^{-*} \textit{and } Y_{CRS}^{O} \xrightarrow{frontier} z^{*} Y_{j} + d^{+*} & \end{array} \qquad \text{(M4.1)}$$

In M4.1, θ^* and z^* represent metric values of the input contraction or output expansion feasible while $s^{-*}, d^{-*}, s^{+*}, d^{+*}$ account for extra gains (slack) for individual input/output variables that are feasible beyond the *pro rata* contraction of inputs or expansion of outputs. The two input/output combinations (X_{CRS}^I, Y_{CRS}^I) and (X_{CRS}^O, Y_{CRS}^O) represent two hypothetical projection points of an inefficient DMU onto the efficient frontier under input contraction and output expansion economic objectives respectively.

The targets estimated in M4.1 can be contrasted with the principles of target setting developed in chapter three. It is argued here that traditional DEA targets fail to accommodate the:

- Value revelation principle as they are strictly input or output oriented,
- Organisational strategy as they do not reflect decision maker preferences,
- Managerial as opposed to strict mathematical feasibility of the estimated targets,
- Global organisational targets as they yield DMU specific targets.

Field and Shutler (1991) appreciate the potential advantages of DEA in a new managerial role, notably target setting. However, they leave this issue to be addressed in future research realising, perhaps, the limitations of the current DEA models to address target setting in its full dimension. Their note on the lack of DEA theoretical studies with target setting orientation, however, perhaps overlooks the work of Thanassoulis and Dyson (1988) who employed target oriented DEA studies (work published in (1992)). Our purpose here is to build on this earlier work of Thanassoulis and Dyson in four respects:

- **First,** to fit the mathematical models developed by Thanassoulis and Dyson (1992) into the managerial context of the principles of target setting justified in chapter three.
- **Second,** to extend and generalise some of the original features of the Thanassoulis and Dyson model concerning the concept of input/output controllability.
- **Third,** to explore the managerial and economic implications of these models.
- **Fourth,** to assess the ability/limitations of these models as tools for resource allocation.

2.2. Models for Prioritised Target (PT) setting

Let us consider the index sets I and O relating to the input ($i=1,...,m$) and output ($r=1,...,s$) variables of assessed DMUs respectively. A partition of the two sets can be made to consider the subset (I_c, O_c) of inputs and outputs that are sought to be improved and the complements (I_f, O_f) not aimed for improvement. The model in M4.2 was developed by Thanassoulis and Dyson (1992) for estimating prioritised targets.

Prioritised target setting (M4.2)	**Estimated targets**
$$\begin{aligned} &\underset{\lambda_j, z_r, \theta_i}{Max} \sum_{r\in O_c} P_r^+ z_r - \sum_{i\in I_c} P_i^- \theta_i + \varepsilon\left(\sum_{r\in O_f} s_r^+ + \sum_{i\in I_f} s_i^-\right) \\ &s.t. \sum_{j=1}^{n}\lambda_j x_{ij} = \theta_i x_{ij_o} \quad i\in I_c \\ &\sum_{j=1}^{n}\lambda_j y_{rj} = z_r y_{rj_o} \quad r\in O_c \\ &\sum_{j=1}^{n}\lambda_j x_{ij} + s_i^- = x_{ij_o} \quad i\in I_f \\ &\sum_{j=1}^{n}\lambda_j y_{rj} - s_r^+ = y_{rj_o} \quad r\in O_f \\ &\theta_i \le 1 \quad i\in I_c \\ &z_r \ge 1 \quad r\in O_c \\ &\lambda_j, s_i^-, s_r^+ \ge 0 \end{aligned}$$	$$\begin{aligned} \hat{x}_{ij_o} &= \theta_i^* x_{ij_o} & i\in I_c \\ \hat{y}_{rj_o} &= z_r^* y_{rj_o} & r\in O_c \\ \hat{x}_{ij_o} &= x_{ij_o} - s_i^{-*} & i\in I_f \\ \hat{y}_{rj_o} &= y_{rj_o} + s_r^{+*} & r\in O_f \end{aligned}$$

Where,

x_{ij}	is the quantity of the i^{th} input of the j^{th} unit,
y_{rj}	is the quantity the r^{th} output of the j^{th} unit,
P_i^-, P_r^+	user specified constants reflecting the decision makers' preferences over the improvement of input/output components,
s_i^-, s_r^+	are slack variables for inputs/outputs non-prioritised to improve,
θ_i, z_r	contraction/expansion metrics ,
ε	this is the Non-Archimedian[1] infinitesimal.

The linear programming model in M4.2 assumes constant returns to scale. The addition of the extra constraint $\sum_{j=1}^{n}\lambda_j = 1$ will mean M4.2 provides targets under the assumption of variable returns to scale (see chapter two). The preferences over inputs/outputs operate as "prices" reflecting the relative importance of different production components in estimating performance targets. **These "prices" reflect the per unit penalty for DMUs that have failed to be efficient in some of the production components respectively.** Higher relative importance for an input (output) will "rotate" the projection of inefficient units to

[1] As explained in chapter two the use of a two phase linear programme to obtain DEA efficiencies can be used in order to avoid the problems encountered with the use of the non-archimedian quantity in the original DEA problem.

those areas of the production possibility set where minimum (maximum) input (output) is desired. As we shall see later the association between the expressed preferences and the assessed targets can be explored using the sensitivity analysis concept in mathematical programming.

Solution to M4.2 yields targets by simultaneous contraction/expansion of controllable inputs/outputs of the production technology. The partitioning of the input/output set into prioritised and non-prioritised classes as proposed by Thanassoulis and Dyson (1992) constitutes an important feature of the PT model. The distinction necessary concerning the prioritisation of inputs/outputs can be relaxed, however, by introducing the notion of the *degree of controllability* that is discussed in more detail later in section 2.3 of the current chapter.

The prioritised target model deviates from the traditional DEA models in two respects:

- *First, it is a target and not efficiency measure seeking method and, therefore, the estimates obtained reflect a focus on individual input/output components.*

The deviation from the radial type of efficiency measures has been advocated long ago by Fare *et al.* (1985), Russell (1985), Charnes *et al.* (1985). However, as mentioned in the last section of chapter two, these studies focus primarily on the ability of non-radial efficiency measures to satisfy properties of their radial counterparts (e.g. homogeneity of degree -1) rather concentrate on the estimation of performance targets.

- *The **second,** deviation concerns the involvement of the decision maker (DM) in the target setting process.*

There is a fundamental difference between the traditional DEA models where no DM involvement is allowed at the target setting process and the PT model where the DM becomes the key component of the target setting process. The aim of this enhancement is twofold: to relax the assumption that target setting is *neutral* in view of the organisational strategy and to extend managerial will and co-operation for accepting the estimated targets. Zeleny (1980) advocating the usefulness of interactive decision analysis methods argues that "*Letting the Man in*", seems to identify a process of considerable promise.

An interactive environment needs to be developed between the analyst and the DM in order to develop preference structures over the desirability of input/output improvements. The formulation of model M4.2 implies that the preference levels over inputs/outputs appear in the objective function simultaneously. Therefore, preferences need to be attached to inputs/outputs in relation to other inputs/outputs. This *relative* basis would accommodate managerial/economic trade-offs in estimating target levels. Mechanisms, such as the *Delphi* method, Khorramshahgol and Moustakis (1988), the *analytic hierarchy process,* Saaty (1980), the *paired comparison* method, Hihn and Johnson (1988) and the *centroid* method, Olson and Dorai (1992) are suggested in the literature as tools to aid management in expressing relative preferences over alternative decisions or sets of criteria. These methods, however, have been tailored for the purpose of multi-objective programming problems and thereby, their applicability in the DEA environment is not guaranteed *per se.*

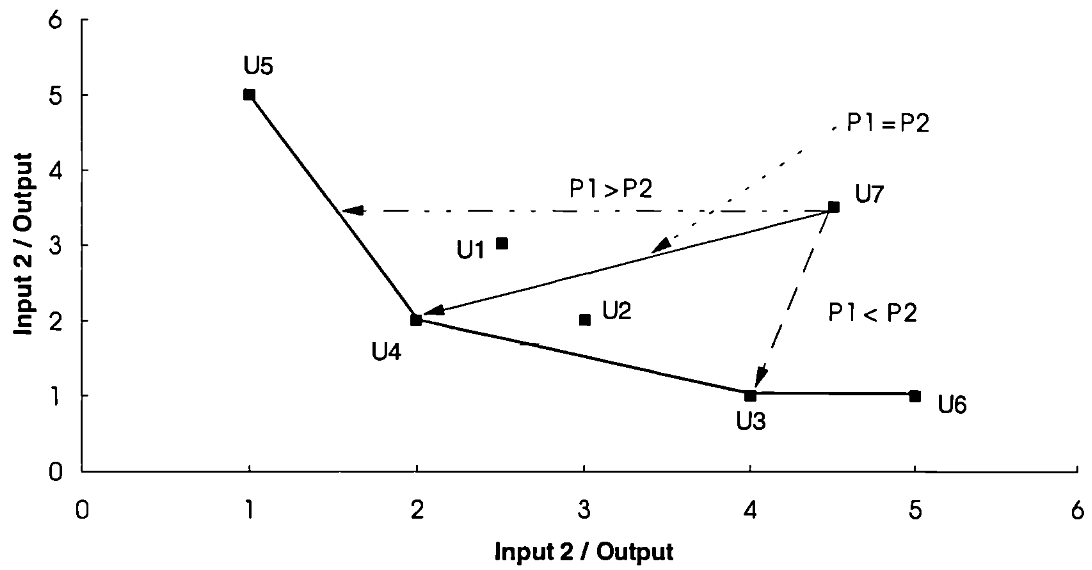
The interactive nature of M4.2 implies that the DMs need to participate and express progressive preferences during the target setting process. It is expected that M4.2 will be used interactively before a satisfactory set of targets is obtained. The design of the target setting process needs to anticipate two important elements.

- The first relates to the amount/nature of information supplied by M4.2 to the decision maker(s) at each phase of the interactive process,
- The second concerns the support provided to decision maker(s) in order to select the most satisfactory sets of targets and terminate the iterative process.

The development of an interactive/iterative target setting framework needs to be customised to the needs and conditions of specific applications. A two dimensional example is employed next to illustrate the features of the PT model. Figure 4.1 exhibits the case of seven operating units and their input requirements for producing one unit of output.

Figure 4.1

A simple example of the Prioritised Target model



Units U5, U4 and U3 are best observed practice units that define an empirical efficient frontier[2] for the set of seven operating units. Using traditional DEA (e.g. M2.8) models the projection of unit U7 will be at unit U4 on the efficient frontier. However, U4 is only one potential efficient projection of unit U7. Had management expressed different preferences over the rate of improvement of the two inputs the projection of U7 would have taken place at different segments along the efficient frontier. Varying the preferences P_i on the desired improvement of one input over the other alternative projections can be obtained on the facet U5U4 with P1 > P2 and unit U3 with P1 < P2.

2.3. Extensions to the basic PT model

In principle, the PT model is the most generalised form by which the non-parametric assessment of performance can be considered. Using the PT model one could generate all known DEA models as special case scenarios. Table 4.1 below contains examples of the association between the PT and some typical DEA models.

---

[2] Unit U6 is located onto the frontier of the production set but not on the efficient subset of this frontier.

Table 4.1
Special cases of the Prioritised Target (PT) model

Prioritised Target model $P_r=P_i=1$	DEA equivalent
$\forall i, \theta_i = 0, z_1 = \ldots = z_m = z$	Output expansion CRS (see chapter two model M2.8)
$\forall r, z_r = 0, \theta_1 = \ldots = \theta_s = \theta$	Input contraction CRS (see chapter two model M2.8)
$\lambda_j \in \{0,1\}$	Free Disposal Hull assumption Athanassopoulos and Storbeck (1992)
$\theta_1 = \ldots = \theta_s = \theta$ $z_1 = \ldots = z_s = z$	Graph Efficiency (Fare *et al.* 1985)

The models exhibited in Table 4.1 are only illustrative; they simply show how the PT can be used to generate known DEA models. However, more extensions can be introduced via the PT model. These extensions are discussed in more detail below.

- ***The first concerns the degree of controllability of inputs/outputs in assessing targets.***

The concept of *degree of controllability* can be introduced to relax the current DEA assumption that inputs/outputs should be considered only as fully controllable or as fully uncontrollable quantities. This new concept can be used to aid target setting in hierarchical (multi-level) organisations. For instance, the labour costs of the branches of a bank are primarily under the control of central management and secondarily under the control of local management. A target setting process for a branch should take into account the differential controllability of labour cost by different levels of management. This can be incorporated in model M4.2 using the following set of additional constraints.

$$L_i \le \theta_i \le 1, \quad L_i \in [0,1], \quad \forall i \in I \qquad \textit{Controllable contraction}$$
$$U_r \le \frac{1}{z_r} \le 1, \quad U_r \in [0,1], \quad \forall r \in O \qquad \textit{Controllable expansion}$$

Where L_i and U_r are lower and upper bounds for the rate of improvement of inputs/outputs correspondingly. Undoubtedly, the values of these bounds are to be decided by the decision makers involved in the assessment of targets. For extreme values of L_i and U_r the model coincides with the PT model in M4.2. For example, $L_i=1$ indicates an uncontrollable input whilst a $L_i=0$ indicates a fully controllable input. The notion of the degree of controllability can be particularly useful when setting targets in multi-level organisations where there are intermediate levels of decision making and no fixed controllability over the inputs/outputs.

Finally, simple sensitivity analysis can assist management in examining how the performance targets vary for different levels of controllability over individual inputs/outputs.

- ***The second extension concerns the restriction currently imposed on the improvement directions taken in projecting inefficient units on the efficient frontier.***

The two constraints $z_r \geq 1$ and $\theta_i \leq 1$ used in M4.2 restrict the directions of projection of inefficient units towards the efficient frontier. The use of a more flexible set of constraints could aid management to discover more suitable targeting strategies for inefficient units. These new targeting strategies have already been introduced in chapter three (section 4) and characterised as *extended Pareto efficiencies.* A revised set of constraints, based on this extension, would enhance the formulation of the PT model in M4.2 to the flexible prioritised target (FPT) model in M4.3.

Flexible prioritised targets (M4.3)

$$\max_{\lambda_j, z_r, \theta_i} \sum_{r \in O} P_r^+ z_r - \sum_{i \in I} P_i^- \theta_i$$

$$\begin{aligned}
s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} = \theta_i x_{ij_o} \qquad i \in I \\
& \sum_{j=1}^{n} \lambda_j y_{rj} = z_r y_{rj_o} \qquad r \in O \\
& L_i \leq \theta_i \leq 1/\Delta_i, \qquad \Delta_i, L_i \in [0,1], \qquad \forall i \in I \\
& U_r \leq 1/z_r \leq 1/\Gamma_r, \qquad \Gamma_r, U_r \in [0,1], \qquad \forall r \in O \\
& \lambda_j \geq 0 \quad \forall j
\end{aligned}$$

The notation in M4.3 is similar to model M4.2 with (L_i , Δ_i) and (Γ_r, U_r) lower and upper bounds for the input/output radial components (Appendix 4B contains a proof that the targets estimated by M4.3 are efficient targets).

With the set of upper and lower bounds used in M4.3 the notion of *degree of controllability* is considered simultaneously with the idea of monitoring the direction of the projections towards the efficient frontier. Each input and output in M4.3 has attached, a user specified degree of controllability, whilst it is allowed to select *extended* orientations for projecting inefficient units on the efficient frontier. **These orientations would allow units to acquire extra resources anticipating higher returns and/or reduce the produced outputs to achieve even higher reductions in the level of resource utilisation.**

The estimation of flexible targets can, however, be complemented further to include conditional constraints that would enhance the selection of directions in the estimation of targets. A plausible assumption, for instance, would be that units acquiring more resource of a given type would need to generate more output of another type. On the other hand, units reducing their activities (reduced output) would also reduce with a higher (perhaps) rate the level of particular inputs used. For example, increasing the number of doctors of a particular specialty in a hospital should increase the number of cases treated related to that specialty. The technical details of incorporating conditions of this type in the flexible targets model, are provided in the appendix for this chapter (Appendix 4C).

3. Setting global performance targets

The importance of global organisational targets was emphasised as one of the principles of effective target setting in chapter three. It was argued that the aggregation of targets estimated for individual DMUs is not sufficient for estimating global performance targets. The assessment of global performance targets as an independent process is important for a number of reasons:

- It develops an operational framework and a basis of planning, independently of targets assessed for individual units,
- It emphasises the interrelations between DMUs from a global perspective,
- It aids developing multi-level target setting processes reflecting the hierarchical structure of MUOs.

The ability of management to set targets at two different levels; one at the DMU level and one at the global level has, undoubtedly, many positive elements. Organisational practice shows that global targets are found almost exclusively in the annual budgeting processes, Richards (1986). Attempts for a more objective assessment of global targets can be found in regulatory bodies e.g. the Audit Commission in the UK where regression analysis models are frequently employed to estimate global targets (e.g. cost savings for local authorities) based on the performance of individual DMUs.

The current methods used for assessing global targets fail to capture the full strength of target setting. The budgeting process, for example, does not contain sufficient systematic elements for estimating performance targets that encapsulate the principles of target setting. Even the use of statistical tools like regression analysis have limited scope as they are mainly driven by cost saving objectives. The method suggested in this thesis, for assessing global targets, has its foundations on the notion of the "industry production function" found in the economic literature.

The assessment of an industry production function has been one of economists' concerns for many years now, Ainger and Chu (1968). Indeed one may refer further back to Farrell's (1957) seminal work where he introduced the notion of *structural* efficiency, an issue that has seen limited expansion in the recent frontier analysis literature. Structural efficiency was introduced by Farrell (1957) in an attempt to assess the performance of industries that cannot be compared with other similar industries due to lack of information and/or homogeneity in operations. For example, it is difficult to compare the performance of the education system (it can loosely defined as an industry) in the UK with another country due to the fundamental differences in the way they are organised. To overcome this difficulty, Farrell argued to compare an industry's performance with an aggregate production function derived from its own constituent firms. **This type of efficiency, namely "structural efficiency", measures the extent to which an industry keeps up with the performance of its own best firms.**

The realisation and definition of the concept of structural efficiency is attributable to Farrell who suggested the use of the arithmetic mean of the technical efficiency of individual firms to estimate the structural efficiency of an industry. This original idea became a more systematic field of study by the work of Aigner and Chu (1968), Forsund and Hjalmarsson (1976), Forsund and Hjalmarsson (1979), Forsund (1992) and Athanassopoulos and Ballantine (1995) developing linear programming models for assessing the structural efficiency of industries.

It is not intended to review the pros and cons of the various industry production function models that can be found in the literature. We shall provide, however, a basic non-parametric formulation inspired by the flexible prioritised target models developed earlier on. Following the same notation used in previous chapters we shall denote a set of j=1,...,n operating units (firms) that use vectors $X \in \Re_+^m$ of input quantities I=1,..i..,m and vectors $Y \in \Re_+^s$ of output quantities O=1,..r...,s. We shall also define two index sets that will separate the inputs and outputs into two classes $I \equiv I_B \cup \bar{I}_B$ and $O \equiv O_B \cup \overline{O}_B$, where I_B and O_B denote index sets of commensurate inputs and outputs. These inputs and outputs will be connected using the set of *balance* constraints in the formulation of the global targets model.

A set of input/output industry specific targets can then be obtained solving the following mathematical programming problem in M4.4. (Notice here that the term "industry" is used interchangeably with the term "global").

Industry based flexible targets **(M4.4)**

$$\underset{\lambda_j, z_r^g, \theta_i^g}{Max} \quad \sum_{r\in O} P_r^+ z_r^g - \sum_{i\in I} P_i^- \theta_i^g$$

$$\begin{aligned} \sum_{j=1}^{n} \lambda_j y_{rj} &= z_r^g \left(\sum_{j=1}^{n} y_{rj} \right) \quad r \in O \\ \sum_{j=1}^{n} \lambda_j x_{ij} &= \theta_i^g \left(\sum_{j=1}^{n} x_{ij} \right) \quad i \in I \end{aligned} \qquad \text{(M4.4a)}$$

$$\begin{aligned} & L_i \le \theta_i^g \le 1/\Delta_i, \quad \Delta_i, L_i \in [0,1], \quad \forall i \in I \\ & U_r \le 1/z_r^g \le 1/\Gamma_r, \quad \Gamma_r, U_r \in [0,1], \quad \forall r \in O \end{aligned} \qquad \text{(M4.4b)}$$

$$\sum_{i\in I_B} \theta_i^g \left(\sum_{j=1}^{n} x_{ij} \right) - \sum_{r\in O_B} z_r^g \left(\sum_{j=1}^{n} y_{rj} \right) \le B \qquad \text{(M4.4c)}$$

$$\lambda_j \ge 0$$

Where,

z_r^g, θ_i^g are global performance metrics for the rth and ith output and input that are sought to be improved respectively,

λ_j are intensity variables that constitute the composite "industry" unit,

B is a user specified fixed term representing the desire to allow violation of the balance constraints between the targets of commensurable inputs-outputs (i.e. $B \in (-\infty, +\infty)$),

P_r^+, P_i^- are user specified preferences over the improvement of inputs and outputs.

The model in M4.4 accommodates multi-input, multi-output production technologies which is an extension of the current econometric literature that uses models with one dependent variable (input or output), Ainger and Chu (1968). M4.4 has clearly a target setting orientation as opposed to an efficiency assessment one. ***The solution of M4.4 involves the estimation of a composite industry input/output mix*** $\left(z_r^g(\sum_{j=1}^{n} y_{rj}), \theta_i^g(\sum_{j=1}^{n} x_{ij}) \right)$ ***derived from some of the best practice units within the industry.*** The linear programming model in M4.4 consists of three sets of constraints that are discussed in more detail next.

- *M4.4a, is used to develop the industry composite unit.*

The estimation of the global performance targets is based on the solution of a single linear programming problem. Following the nature of typical DEA models a composite unit is developed that seeks to outperform the assessed unit. As the assessed unit $\left(\sum_{j=1}^{n} x_{ij}, i \in I; \sum_{j=1}^{n} y_{rj}, r \in O\right)$ is the whole industry or organisation, its corresponding composite unit $\left(\sum_{j=1}^{n} \lambda_j^* x_{ij}, i \in I; \sum_{j=1}^{n} \lambda_j^* y_{rj}, r \in O\right)$ yields input-output targets for the whole organisation. Targets corresponding to controllable inputs and outputs provide information on the highest expected global performance of the organisation given the preference structure reflected in the objective function of M4.4 for input/output improvements.

- *M4.4b is used to facilitate the flexible target setting policies introduced in M4.3.*

The estimation of global organisational targets is affected by the upper and lower bounds set over the expected improvements on the inputs and outputs. These upper and lower bounds can be used to facilitate the development of planning scenarios regarding the global organisational achievements.

- *M4.4c represents the so-called balance constraints which seek to monitor the relation between the aggregate target levels of commensurate inputs/outputs.*

Consider for example, a target setting process for local governments with revenue sources as input factors and investments and services provided as outputs factors. At the global level one would expect local authorities (industry) to balance revenue sources (e.g. governmental and own) with expenditures. Moreover, these constraints give the opportunity for developing financial scenarios that include (if necessary) short run deficit/profit levels at the aggregate industry level. Balance constraints can also be used in the estimation of DMU specific targets but their presence is more important in the estimation of the "industry" based targets.

- *The form of the objective function of M4.4 affects the selection of efficient units in the composite industry unit.*

The composite unit used for assessing the industry based targets constitutes one particular facet of the production possibility set. There are a number of alternative facets that could be used as the basis of the composite industry unit. The selection of any particular facet is based on the form of the objective function of M4.4 and the magnitude of P_r and P_i

reflecting preferences over individual inputs/outputs. The selection of a particular facet for deriving the industry based targets would yield corresponding marginal productivities between the inputs and outputs. The set of marginal productivities would characterise the efficient tradeoffs between inputs/outputs across the industry concerned. Assessing corporate performance of the grocery industry, for example, one can compare the substitution between capital employed and labour expenses estimated by M4.4 to summarise the state of technology at the corporate industry level. These substitutions can then be compared with those substitutions obtained by ordinary DEA for individual grocery firms and identify those firms that specialise in a particular operation and those that behave closely to the general industry mode.

The industry production function approach yields aggregate input/output targets. These targets provide useful information of the potential gains and savings that the industry could, in principle, achieve. It is noteworthy, however, that the assessment of industry based performance targets can over or under estimate the true potential of the assessed industry as they are based on the best <u>observed</u> performance within the industry concerned. The structural targets in an industry with overall poor performance, for example, would yield moderate targets which would underestimate the true potential of the sector.

Furthermore, when it comes to identify the effects of the aggregate targets on individual units the model M4.4 lacks the ability to focus on the targets of individual units. This has serious drawbacks in the resource planning and goal setting processes in organisations. The lack of *decomposition,* of the global targets assessed among individual units, constrains the ability of the industry target model to become an effective tool for supporting resource allocation decisions in MUOs. This issue, however, will be resolved with the centralised and decentralised target-based planning models that will be developed in chapters five and six.

The preceding two sections emphasised the development of appropriate target setting models for supporting the principles of effective target setting introduced in chapter three. The prioritised target model developed by Thanassoulis and Dyson (1992) were extended to the flexible and industry target setting models as a means for addressing the principles of target setting operationally. The next section focuses on the economic and managerial implications of the models developed earlier. It is anticipated that this discussion will highlight the decision support character of the target setting models towards resource allocation decisions.

4. Economic information from the prioritised target models

The prioritised target model in M4.2 was developed in an attempt to estimate targets compatible with the principles of target setting put forward in chapter three. The managerial and economic implications of this model are examined in the remainder of this chapter. This information will be used as the basis for developing decision support scenarios in chapter six.

4.1. DEA & marginal rates of transformation: Background

Koopmans ((1951) and (1957)) emphasised the importance of the concept of substitutability in economics and management science. Substitutability or marginal rates of transformation (MRT) are concerned with the extent to which simultaneous changes between factors of production can occur without, however, affecting input/output factors not involved in the substitution. The definition of MRT is conditioned by the notion of optimality concerning the quantities of inputs used for delivering maximum amounts of outputs, (i.e. production function). In classical economics, production functions are usually defined having continuous derivatives throughout their domain (e.g. the Cobb-Douglas function). In this case the MRT are uniquely defined at all points in the interior of that domain and are estimated using econometric methods, Greene (1990).

The production function in DEA has a piecewise parametric form which implies that the efficient frontier is made of linear segments which in the general case are supporting hyperplanes[3]. Each assessed unit c lies on a supporting hyperplane SH^c which is made of the efficient units (X^j, Y^j) that satisfy the following mathematical relation (under variable returns to scale):

$$SH^c = \left\{ (X^j, Y^j) \,\middle|\, \mathbf{u}^c Y^j - \mathbf{v}^c X^j - \omega^c = 0 \right\}.$$

Where,

$\mathbf{u}^c \in \Re^s$ and $\mathbf{v}^c \in \Re^m$ are respectively vectors of optimal weights attached to outputs and inputs of the assessed unit c. These vectors constitute the *norm* of the supporting hyperplane,

$Y^j \in \Re^s_+$ and $X^j \in \Re^m_+$ are respectively the output and input vector of DMU j.

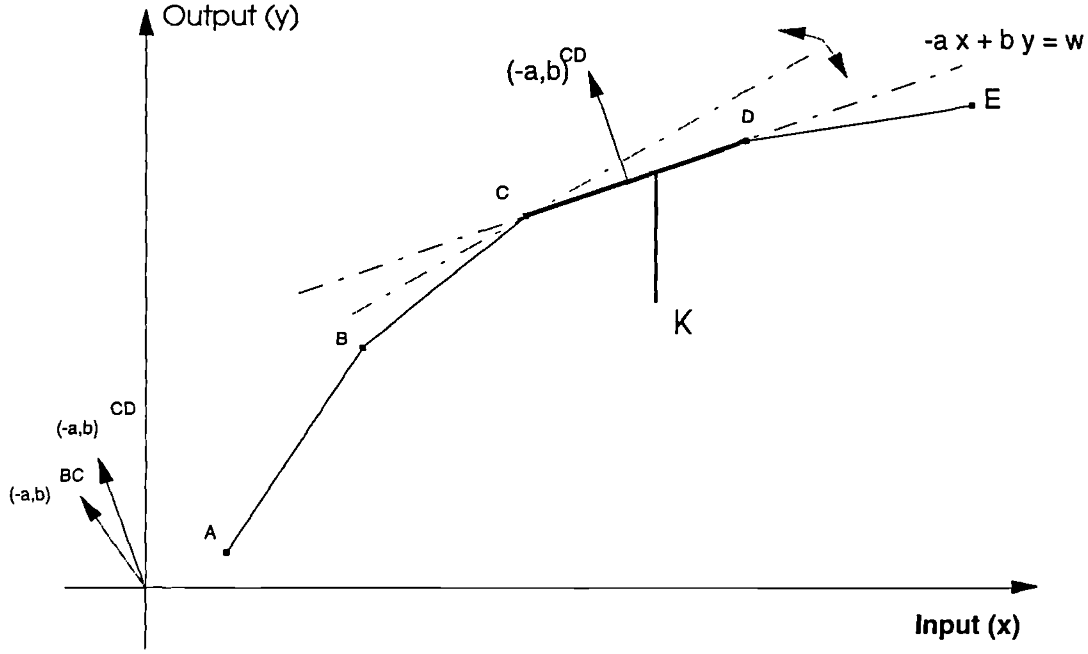
ω^c is a factor reflecting the nature of the returns to scale at unit c (see section 3.6 of chapter two).

---

[3] A supporting hyperplane H can be defined as follows: H={x | $a^T x = 0$}, where a is the norm of the hyperplane.

The number of efficient (and non-collinear) DMUs that satisfy the conditions of a particular supporting hyperplane constitute its dimension $|SH^c|$. The dimension of a supporting hyperplane will be connected later to the definition of MRT between the inputs and outputs. A simple input/output production technology made of units A to E in Figure 4.2 is used next to demonstrate the concept of efficient facets and MRT in DEA. The discussion will focus on the assessment on the inefficient unit K.

Figure 4.2
Efficient facets and rates of substitution



The efficient frontier of the production possibility set in Figure 4.2 is segmented into facets defined by efficient observed input-output points. The efficiency of unit K is assessed using the supporting hyperplane defined by units C and D which can be expressed mathematically using the equation $SH^K \equiv \{(x^j, y^j) | by^j - ax^j - w = 0, j = C, D\}$ where (-a, b) define the *norm* to the hyperplane. The MRT of input and output combinations between point C and D are obtained by the ratio of partial derivatives with respect to input x and output y. This can be stated mathematically as:

$$MRT_{x,y} = \frac{\partial(by - ax - w)/\partial y}{\partial(by - ax - w)/\partial x} = -b/a$$

Inefficient units projected on the segment CD will be characterised by the MRT of their supporting hyperplane (-b/a). Each segment, AB, BC, etc. in Figure 4.2 yields a different set of MRT. **Note however, that extreme points (units) of an efficient frontier (such as**

unit C) are supported by an infinite number of hyperplanes which implies a non-uniqueness of MRT for these particular points. In our case supporting hyperplanes defined by a single efficient unit, say C, do not define a unique set of MRT whilst in the case of supporting hyperplanes defined by two extreme[4] units of the frontier there is a unique set of MRT.

The example in Figure 4.2 shows the association between the number of efficient units that are used to define a supporting hyperplane and the extent to which the MRT of this hyperplane are uniquely defined (see Bessent *et al.* (1988)).

In the general case, however, the definition of supporting hyperplanes in DEA is made via the solution of linear programming problems as introduced in chapter two (see model M2.11). Let us consider the case of the constant returns to scale efficiency model reproduced in M4.5.

Peer sets and marginal rates of substitution (M4.5)	
Offensive model	**Defensive model**
$\underset{\lambda_j, z}{\text{Max}} \quad z$ $-\sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- = -x_{ij_o} \quad i = 1,\ldots,m$ $\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = z y_{rj_o} \quad r = 1,\ldots,s$ z *free* and $\lambda_j \geq 0, \forall j \;\; s_i^-, s_r^+ \geq 0$	$\underset{v_i, u_r}{Min} \quad \sum_{i=1}^{m} v_i x_{ij_o}$ $s.t. \quad \sum_{r=1}^{s} u_r y_{rj_o} = 1$ $\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} - t_j = 0$ $v_i, u_r, t_j \geq 0$
Strong Complementary Slackness Condition, Spivey and Thrall (1970)	
$\lambda_j^* \times t_j^* = 0 \qquad \lambda_j^* + t_j^* > 0$ $s_i^{-*} \times v_i^* = 0 \quad and \quad s_i^{-*} + v_i^* > 0$ $s_r^{+*} \times u_r^* = 0 \qquad s_r^{+*} + u_r^* > 0$	

The LP models in M4.5 have a dual relation and therefore their solutions are linked via the strong complementary slackness condition (SCSC). The implications of this condition in a DEA context are discussed next.

[4] Extreme units (vectors) of a frontier are linearly independent (non-collinear) vectors, Hadley (1961).

The number of basic, λ, variables in the optimal solution of the offensive model in M4.5 can be m+s-1 where m and s are the number of inputs and outputs in the model. The exact number, however, is traded between the λ_j, s_i^-, s_r^+ variables, where $\lambda_j^* > 0$ identifies[5] efficient units used to define the supporting hyperplane of an assessed unit j_O. The number of units with, $\lambda_j^* > 0$, can vary from 1 to m+s-1 which implies, however, that the number of positive slacks (s_i^{-*}, s_r^{+*}) can also vary from 0 to m+s-2 in the optimal solution. As the slacks are associated via the SCSC (see M4.5) with the weights (v_i^*, u_r^*) in the offensive DEA model we can summarise as follows:

> **The variation in the number of efficient units included as peers ($\lambda_j^* > 0$) in the optimal solution of the DEA model in M4.5 affects the number of inputs/outputs that have positive weights (v_i^*, u_r^*) in the solution of the defensive formulation of the same problem.**

These implications are discussed next.

- Inefficient units with supporting hyperplanes that include less than m+s-1 efficient units ($\lambda_j^* > 0$) have positive (basic) values in some of the input/output slack variables. The positive slacks imply zero weights for the corresponding input/output due to the SCSC.
  (I.e. if $\exists i\!: s_i^* > 0 \Rightarrow v_i^* = 0$ and if $\exists r\!: s_r^* > 0 \Rightarrow u_r^* = 0$).

- The MRT between inputs/outputs of a DMU j are defined as ratios of the weights attached to inputs/outputs in the solution of the basic DEA model in M4.5. For an efficient unit j_O the following relation holds between its inputs/outputs: $\sum_{i=1}^{m} v_i^* x_{ij_o} - \sum_{r=1}^{s} u_r^* y_{rj_o} = 0$. The marginal rates of transformation between input i and output r that would allow unit j_o to remain efficient can be expressed mathematically as $MRT_{x_i, y_r} = \dfrac{\partial x_i}{\partial y_r} = -{u_r^*}/{v_i^*}$.

- Efficient hyperplanes that are defined by less than m+s-1 efficient units imply that some of the inputs and/or outputs will have zero or infinite MRT, Charnes *et al* (1978) and Thomson *et al* (1990).

---

[5] Units with $\lambda_j^* > 0$ in the offensive model have zero t_j value in the defensive model which implies that they are used to define the supporting hyperplane $SH^{j_o} = \left\{ (X^j, Y^j) \,\middle|\, \mathbf{u}^{j_o} Y^j - \mathbf{v}^{j_o} X^j = 0 \right\}$ for unit j_O under constant returns to scale.

The association between marginal productivities and DEA has not been neglected by the current DEA literature. One can distinguish two DEA research streams concerned with the identification of supporting hyperplanes for inefficient units (*envelopment* approach) and the estimation of MRT between the inputs and outputs (*value-based* approach). As both streams cover a very wide span this thesis discusses their rationale and some of their strengths and weaknesses avoiding technical details.

- ***The envelopment approach***

The *envelopment approach* seeks to address the problem of projecting inefficient units on efficient facets of less than full dimension, (i.e. less than m+s-1 peers assuming constant returns to scale). As the efficient facets obtained by ordinary DEA models may not have full dimension, the *envelopment* approach projects inefficient units on *extensions of full dimensional facets* that fall outside the production possibility set. The estimation of MRT between inputs/outputs is a natural by-product of the existence of full dimensional facets. The rationale of the methodology is illustrated further using the two input per unit of output diagram in Figure 4.3.

Figure 4.3

Envelopment of inefficient units on full dimension facets

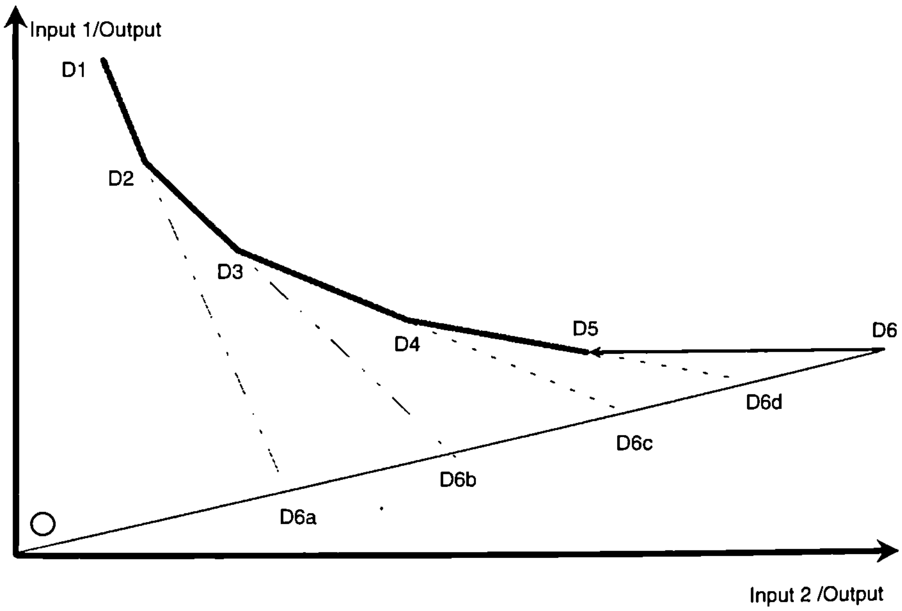


Figure 4.3 contains a set of five efficient units D1,D2,D3,D4 and D5 whilst unit D6 is an inefficient unit as it uses the same quantity of input 1 but more of input 2 per unit of output compared with DMU D5. Using an ordinary DEA assessment unit D5 is the peer for unit D6 and, therefore, its targets will be equivalent to the input/output values of unit D5. An

improvement for Input 2 will be estimated as a positive slack which is equivalent to the segment D5D6. The positive slack implies that the weight associated with input 2 will be zero and, therefore, there are no MRT defined between input 1 and input 2. The MRT for unit D6 will be defined properly only when the unit is projected along one of the *extensions* of the facets D1D2 or D2D3 or D3D4 and D4D5. The efficiency of unit D6 will be modified according to the facet selected for projection (e.g. OD6a/OD6 when compared with the D1D2 efficient facet). On the targets estimation side, however, unit D6 will be projected outside the empirical frontier of the production possibility set and this is one of the main limitations of these methodologies.

A variety of technical models can be found in the literature for projecting inefficient units on *full dimensional facets*. Bessent *et al.* (1988), Clarke (1988) developed the so-called *constraint facet analysis*, whilst Olesen and Petersen (1991), (1993) developed mixed-integer DEA models for projecting inefficient units on full dimensional facets.

- ***The value based approach***

The *value-based approach* seeks to assess efficiency by superimposing "values" on the inputs and outputs in a DEA assessment. As the assessed weights are bounded within user-specified regions it is obvious that the DEA solution will attach positive weights to those inputs/outputs that are restricted. These models constitute the weights restrictions research area in DEA. A variety of different models have been developed by Dyson and Thanassoulis (1988), Charnes *et al.* (1990), and Oral *et al.* (1991). The estimated positive weights on the inputs/outputs are not in principle associated with the estimation of efficient facets of full dimension. However, recent research lead by Dyson *et al.* (1993) showed that the use of weight restrictions leads to revised efficient frontiers (*value frontiers)* in order to satisfy these restrictions. A *value frontier* can be defined by supporting hyperplanes different to those defined by ordinary DEA analysis.

In summary, the two research streams provide very useful insights into the problem of estimating supporting hyperplanes and weight multipliers for inefficient DMUs. It is argued, however, that both methods cannot support reliably the problem that this research seeks to address, namely how best to utilise the principles of target setting for decision making purposes. The envelopment approach attempts to extrapolate units on "full dimension facets" outside the observed efficient frontier and, therefore, the corresponding targets, yield excessive input/output improvements. This can endanger the principle of *managerial feasibility* of the assessed targets. On the other hand, the "value based" approach seeks to assess relative efficiency and not performance targets, and also redefines the DEA frontier as defined in chapter two (see model M2.11) using judgmental *value frontiers*.

4.2. Prioritised targets and productivity tradeoffs

As already discussed, the prioritised target model in M4.2 deviates from the radial DEA models developed originally by Charnes *et al.* (1978), and Banker *et al.* (1984). It also deviates from Charnes *et al.* (1985c) attempts to introduce non-radial models. One of the unique characteristics of the prioritised target model concerns the incorporation of decision making preferences during the target setting process. The implications of this feature on the estimation of marginal rates of transformation between inputs/outputs is examined next. To facilitate this discussion the dual formulation of M4.2 is used as stated in M4.6.

Weights based prioritised targets model (M4.6)

$$\min_{\alpha_r,\beta_i,\gamma_r,\zeta_i} \sum_{i\in I_c}\beta_i x_{ij_o} + \sum_{i\in I_f}\zeta_i x_{ij_o} - \sum_{r\in O_c}\alpha_r y_{rj_o} - \sum_{r\in O_f}\gamma_r y_{rj_o} + \sum_{r\in O_c}P_r^+ - \sum_{i\in I_c}P_i^-$$

subject to

$$\sum_{i\in I_c}\beta_i x_{ij} + \sum_{i\in I_f}\zeta_i x_{ij} - \sum_{r\in O_c}\alpha_r y_{rj} - \sum_{r\in O_f}\gamma_r y_{rj} \geq 0 \quad (\forall j=1,\dots,n)$$

$$\alpha_r y_{rj_o} \geq P_r^+ \quad (\forall r\in O_c)$$

$$\beta_i x_{ij_o} \geq P_i^- \quad (\forall i\in I_c)$$

$\alpha_r, r\in O_c; \beta_i, i\in I_c$ are free variables;

$\gamma_r \geq \varepsilon, r\in O_f; \zeta_i \geq \varepsilon, i\in I_f$, and $0<\varepsilon \lll 1$.

Where,

x_{ij}	is the quantity of i^{th} input of the j^{th} DMU,
y_{rj}	is the quantity of r^{th} output of the j^{th} DMU,
P_i^-, P_r^+	are user specified preferences for input/output improvements.

The weights based prioritised target model reproduced in M4.6 seeks to minimise the weighted sum of the differences between the inputs and outputs of DMUs. The weights attached to controllable inputs/outputs are unrestricted variables, whilst a weak positivity (ε) assumption holds for the weights attached to uncontrollable inputs/outputs.

The first system of constraints expresses the fundamental input-output equations that determine the mechanisms for assessing performance: **no unit can generate a weighted sum of outputs greater than its corresponding weighted sum of inputs.** The second and third system of constraints relate to the weighted value of each controllable input and output of the assessed DMU. These constraints force the weighted values to be greater or equal to the values used to capture the DM's preferences over the improvement of the prioritised inputs/outputs in assessing targets. A similarity exists between the weight constraints associated with the inputs/outputs of DMUs assessed by model M4.6 and those of the value frontiers' literature discussed earlier.

The solution obtained by model M4.6 has important implications concerning the estimation of marginal rates of transformation discussed earlier. The structure of model M4.2 and its dual M4.6 guarantee positive values for all weights α_r, β_i that concern the outputs and inputs that are to be improved. This statement can also be extended to the case of inputs/outputs that management has partial degree of controllability addressed by the flexible prioritised targets model in M4.3. (The dual formulation of M4.3 and the relevant discussion are provided in Appendix 4A).

The solution of M4.6 does not always give non-zero weight values (ζ_i, γ_r) for inputs and outputs that are not prioritised to be improved. As discussed earlier, zero weights in DEA imply positive slack values for some inputs and outputs involved in the assessment. This in turn implies that the supporting hyperplanes obtained by M4.6 do not have full dimension.

In summary, the positive weights obtained from M4.6 can be interpreted as marginal rates of transformation. The possibility of considering non-fully defined MRT in empirically defined production functions was not ruled out by Koopmans (1957).

> "The price system associated with an efficient point likewise represents the more general concept, which can serve to determine marginal rates of substitution whenever the latter are definable, but which remains available as an instrument of decentralisation of decisions even if marginal rates of substitution do not exist or depend on the directions of net output change considered". (pp. 95)

The use of linear programming models for assessing performance targets is associated with the so-called post-optimality analysis, which provides information related to changes in the optimal solution due to marginal changes in the prices/costs of the objective function. The economic implications of sensitivity analysis in linear programming have been emphasised by Charnes and Cooper (1961). In a later attempt they focused on the differences between economics and operational research in the treatment of the concept of marginal rates of transformation, Charnes and Cooper (1980).

4.2.1. MRT and imputed prices in linear programming

The solution of a linear programming problem yields information on the extent to which its optimal solution remains unchanged for marginal changes of the objective function coefficients. Using sensitivity analysis information, obtained from the optimal solution of the linear programming problems in M4.2 and M4.3, a number of observations can be made.

- *The imputed (shadow) prices of the inputs/outputs in the solution of M4.2 and M4.3 can be interpreted as the marginal rates of transformation between inputs/outputs.*
- *The estimated imputed prices are affected by the preference structure of decision makers over the improvement of the inputs/outputs in M4.2 and M4.3. In other words, different sets of preferences may yield a different set of imputed prices and thus MRT for an assessed DMU.*
- *Using sensitivity analysis one can obtain information on how imputed prices change by varying the preference levels on the inputs/outputs of the objective function of M4.2 and M4.3. Simultaneous variation of the preferences on more than one input/output can be addressed using parametric programming.*

Estimation of MRT using models M4.2 and M4.3 is not based on the development of supporting hyperplanes of full dimension. As mentioned earlier, this may require the extension of the efficient frontier outside the "natural" frontier of a production possibility set which is outside the scope of this thesis. The imputed prices derived by M4.2 and M4.3, on the other hand, are based on the DEA frontier of the assessed units and, therefore, cannot be interpreted as MRT in the way they are conceived in economic theory[6]. This does not prevent, however, their use for supporting marginal analysis and resource allocation in the spirit of Koopmans (1957). The use of the MRT for decision support purposes is discussed in more detail in chapter six of the thesis.

4.3. An illustrative example

A three output one input production technology will be used next to illustrate the economic information obtained from the prioritised target setting models. Table 4.3 exhibits the output values for eight operating units used in the example. The numerical illustration has three stages. In stage one the prioritised target model M4.2 is applied to assess targets for unit K. Secondly, the assessment of marginal productivities for relatively efficient units is

---

6 Recall here that in economic theory RT are well defined as partial derivatives of continuous and differentiable production/cost functions.

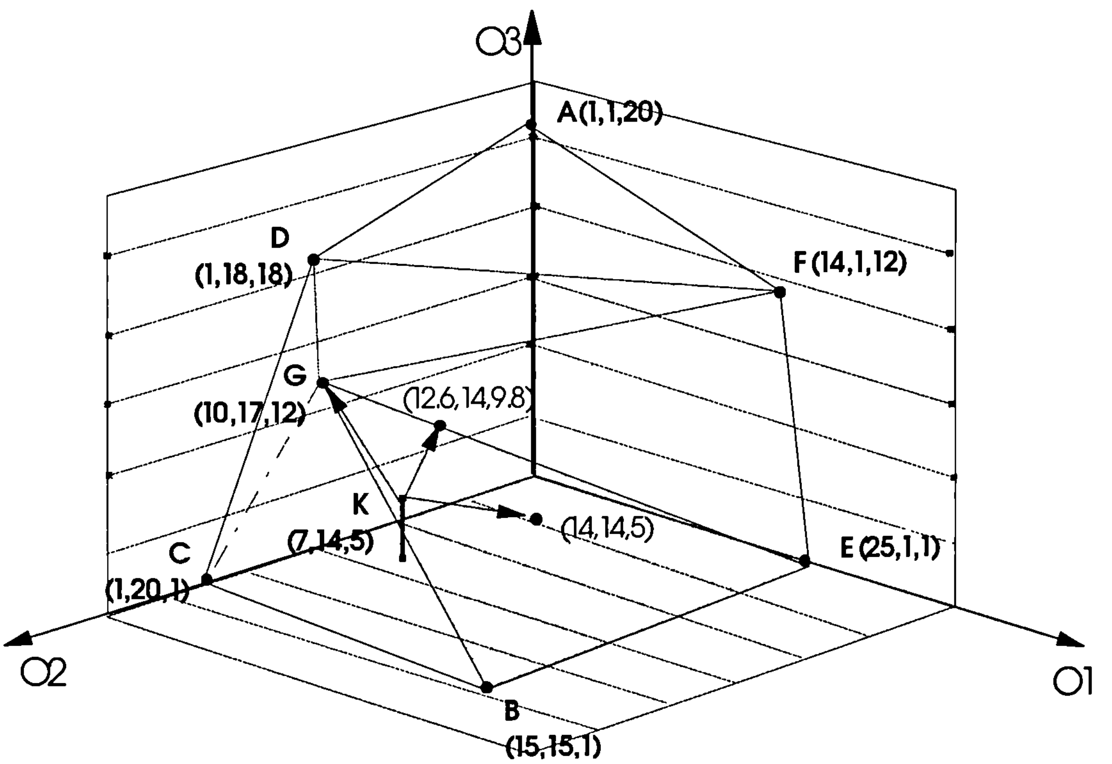
discussed using unit G as an example. Finally, the assessment of targets and marginal productivities using the global targets model M4.4 is discussed in comparison with the previous two stages. Results from the first stage are included and discussed in the main body of this chapter. The remaining two stages are discussed in the appendix of this chapter (Appendix 4D and 4E).

Table 4.2
A three-output one-input production technology

Unit	A	B	C	D	E	F	G	K
Output 1	1	15	1	1	25	12	10	7
Output 2	1	15	20	18	1	1	17	14
Output 3	20	1	1	18	1	14	12	5
Input	1	1	1	1	1	1	1	1

Prior to applying the various target setting models a discussion on the nature of the efficient frontier and its efficient facets will be made. It can be seen in Figure 4.4 that in a three output one input production technology an efficient facet contains three efficient (non-collinear) units in order to have full dimension.

Figure 4.4
Prioritised Targets and Productivity Tradeoffs



Let us consider the efficient hyperplane BGE. To define this hyperplane the coordinates of three non-collinear points are required. In DEA, this is feasible when three efficient units are present in the optimal basis of the corresponding linear programs. An efficient facet with full dimension has well defined and continuous partial derivatives with respect to inputs/outputs of DMUs in the interior of the efficient facet (supporting hyperplane). One needs to examine, however, the case of having inefficient units projected on non-full dimensional facets and the estimation of MRT in these cases.

Projections of inefficient units onto facets with less than full dimension imply that marginal productivities cannot be defined in full accordance with the economic definition of substitutability. Inefficient units projected on the line GE, for instance, will have well defined productivity tradeoffs between output O1 and O2 but there will always be a fixed proportions[7] between these two outputs and O3. The *radial* type of DEA models, (CRS/VRS) will give zero or infinity productivity tradeoffs for output O3 while the *nonradial* models yield positive productivity tradeoffs (in fixed proportions).

- **Resource allocation implications of projections above/below a full dimensional facet**

The polyhedron BGFE is made of the two supporting hyperplanes GEF and GEB which are full dimensional facets of the frontier. Two half-spaces with different economic implications can be defined using the polyhedron BGFE. The space in the interior of BGFE is dominated by the efficient frontier of the example. Input/output sets within this space will, therefore, be inefficient. On the other hand the space outside the intersection of the two half spaces (GEF and GEB) lies outside the current efficient frontier. Input/output sets within this space are not feasible in view of the boundaries of the efficient frontier.

These two observations indicate the considerations that need to be made in using exchange rates for resource allocation. *A resource allocation policy, therefore, even when it is based on efficient exchange rates between inputs/outputs can guarantee the individual units' efficiency for only a limited range of input/output combinations. However, as we shall see in the following chapters, efficiency is only one of three basic objectives, namely efficiency, effectiveness and equity, in a performance based resource planning process.*

---

[7] These fixed tradeoffs have similarities with the Leontief type production functions of a fixed technology.

- **Projection on a non full-dimensional facet (e.g. segment GE)**

Our previous point regarding the links between resource allocation and MRT will be illustrated better focusing on the non-full dimensional facet GE. Out of the infinite number of hyperplanes that pass through the segment GE the subset that spans the half-space outside the envelope BGE-GEF will always lead to input/output combinations that lie on or dominate the current efficient frontier. On the other hand, the subset of hyperplanes passing through GE that span the inner part of the envelope BGE-GEF could lead into input/output combinations that are dominated by the current efficient frontier. As mentioned earlier, the use of the "efficient" exchange rates in a resource allocation process will need to be supported appropriately so that inefficient input/output combinations can be avoided.

- **Sensitivity analysis and marginal rates of transformation**

We shall now turn our attention on the assessment of performance targets for unit K. The use of decision makers' preferences for input/output improvements required by M4.2 was organised as follows: We start from a "neutral" or indifferent initial case where all outputs are given an equal priority of unity. The optimal solution of this problem yields information regarding the extent to which the preferences, over the improvement of the three outputs, can change without affecting the selection of efficient facet of projection. This strategy can help decision makers to explore the consequences of alternative preference structures on the selection of efficient peers, the estimated targets and the exchange rates between the inputs and outputs. The LP problem solved in the example was derived from M4.2 as follows.

$$\max_{z_1,z_2,z_3,\lambda_j} \quad P_1^+ z_1 + P_2^+ z_2 + P_3^+ z_3$$

$$\sum_{j=1}^{8} y_{1j}\lambda_j = 7z_1$$

$$\sum_{j=1}^{8} y_{2j}\lambda_j = 14z_2$$

$$\sum_{j=1}^{8} y_{3j}\lambda_j = 5z_3$$

$$\sum_{j=1}^{8} x_j\lambda_j \le 1$$

$$z_1, z_2, z_3 \ge 1 \text{ and } \lambda_j \ge 0$$

Table 4.3 shows results obtained by the sequential variation of preferences in M4.2.

Table 4.3
Prioritised targets & exchange rates for unit K

Priority Levels for Efficiency Components (allowable increase-decrease)				Productivity Tradeoffs			
Output selected	P_1^+ (Incr., Decr.)	P_2^+ (Incr.,Decr.)	P_3^+ (Incr., Decr.)	α_1/α_2	α_1/α_3	α_2/α_3	FACET
O1,O2,O3	1 (+0.68, -0.11)	1 (+2.2, -0.43)	1 (+0.13, -0.61)	1.992	0.714	0.35	G
O1	**1.7** (+2.88, -0.01)	1 (+0.023, -∞)	1 (+0.012, -0.6)	3.3218	1.214	0.3655	EG
O1	**4.7** (∞, -0.118)	1 (+5.26, -∞)	1 (+0.025, -∞)	1.5	3.27	2.18	EGB*
O2	1 (+0.015, -∞)	**3.5** (+21.7, -0.3)	1 (+5.56, -0.01)	0.58	0.725	1.25	GD
O2	1 (+0.11, -∞)	**25.3** (∞, -1.25)	1 (+28.5, -∞)	0.16	1.51	9	GC
O3	1 (+1.11, -∞)	1 (+0.3, -∞)	**6.8** (∞, -0.565)	3.25	0.619	0.1838	FGD*

The first column in Table 4.3 indicates the output selected for examining the implications of sensitivity analysis information on the estimation of MRT. For example, in the first row the preferences are equal among all outputs ($P_1=P_2=P_3=1$). In the optimal solution, each output has an allowable range that its preference can vary without altering the optimal solution. The preference (P_1) for the first output in the first row can increase from 1 to 1.68 and decrease to 0.89 before a new optimal solution is obtained. The revised optimal solutions and their implications are shown in the remaining rows. The numbers in bold typeface indicate the outputs whose preference is progressively strengthened. The actual magnitude of the preference levels are obtained using sensitivity analysis from the solution to model M4.2.

The productivity tradeoffs for alternative preference structures are listed in columns five to seven. These values were obtained from the ratios of the estimated weights ($\alpha_1,\alpha_2,\alpha_3$) of outputs (O1,O2,O3) respectively (defined in model M4.6). The inputs and outputs of an efficient DMU j_O included in our example need to satisfy the following relationship: $\alpha_1^* y_{1j_o} + \alpha_2^* y_{2j_o} + \alpha_3^* y_{3j_o} = 1$. Unit j_O will remain efficient for all output combinations that would not violate the previous relationship. This implies that an increase of output O2 by one unit would result to a decrease to output O1 by α_2^*/α_1^* units and keeping the quantity of O3 unchanged.

The last column lists the efficient facets for unit K as they change by varying the preferences over individual outputs. **These efficient facets can be used to provide a finite set of alternative options to decision makers facilitating the interactive process for setting targets compatible with their preferences.**

Six alternative preference structures have been applied in assessing performance targets for unit K. At each step the preferences for one particular output were adjusted in line with the sensitivity analysis information so that the facet of projection changes. This was repeated for each output until no further changes in the optimal solution (or factor projections) could be obtained by increasing the preference on a particular output. Some observations emanating from the results obtained in Table 4.3 are discussed below.

- *There is an association between the projection of inefficient units to an efficient facet and the preferences expressed over input/output improvements.*

Small changes in the relative priorities result in big changes insofar as the projection facets and the corresponding targets are concerned. For example, when the preference level for output O1 is changed from 1.7 to 4.7 the implications for the estimated productivity tradeoffs were surprising. Higher preference for a production component would "force" the efficient projection to take place to an area of the frontier where the corresponding component has high substitutions with the other inputs/outputs.

In the first case, when the preference was 1.7 for O1 and 1 for O2 the trade-off between outputs O1 and O2 was 3.32, i.e. for an extra unit of output O1 we need to give up 3.32 units of output O2. However, a stronger preference value over output O1 (4.7 in row three from 1.7 in row two) resulted in weaker substitution between output O1 and O2. In particular, for each extra unit of output O1 produced we needed to give up only 1.5 units of output O2. The revised projection of unit K has taken place in an area of the efficient frontier that the production level of output O2 is not substantial.

On the other hand, the productivity tradeoffs between output O1 and output O3 were strengthened (from 1.214 to 3.27) as the preference over output O1 changed from 1.7 to 4.7. Evidently the increased preferences over the improvement of output O1 has projected the inefficient unit on a facet of the frontier that output O2 has also strong substitutions with output O1, and *all of this at the expense of output O3* which has weakened its substitutions with output O1 further from stage 1 to stage 2.

- *A new instrument for linking the preference structure with the selection of efficient facets is used, that is sensitivity analysis at the optimal solution of M4.2.*

Undoubtedly, the most important issue arising from the preferences' variation are the changes to the efficient facets to which inefficient units are projected. It must be noticed, however, that the potential benefits of sensitivity analysis information have not been explored in full in our example. For instance, parametric programming can be combined with the selection of appropriate preferences and, thus, simultaneous variation on input/output preferences can be employed.

- *There are resource allocation and production implications that follow any decisions to project inefficient units onto different facets of the efficient frontier.*

The selection of efficient facets for projecting inefficient units has significant implications on the input requirements for delivering outputs. To illustrate this issue we shall consider the full dimensional facets EGB and FGD from the above numerical example. The MRT that apply along the two facets were estimated earlier as:

Facet	α_1/α_2	α_1/α_3	α_2/α_3
EGB	1.5	3.27	2.18
FGD	3.25	0.69	0.18

Production of the three outputs along the efficient facet EGB yields high substitutions between output O3 and the other two outputs (i.e. you need to give up 3.27 or 2.18 units of output O3 for an extra unit of output O1 or output O2 respectively). Moreover, each extra unit of output O1 will cause the loss of 1.5 units of output O2. In summary, projection of unit K onto the facet EGB implies that the decision maker seeks to increase production of output O1 primarily and output O2 secondarily at the expense of producing output O3.

A different picture holds for units projected on the facet FGD. The substitution between O1 and O2 has changed from 1.5 in facet EGB to 3.25 which implies that an extra unit of O1 can be produced by giving up 3.25 units of O2. Secondly, the production of output O3 is economically more attractive in facet FGD than in the facet EGB. Indeed, an extra unit of output O1 means giving up 0.69 units output O3 or an extra unit of output O2 means giving up only 0.18 units of output O3 respectively for the DMU to remain on that facet.

The information obtained by investigating the economic implications of alternative efficient facets can be used as the basis for selecting the most appropriate set of targets for inefficient units. A finite number of projection facets can be obtained for each inefficient unit using the sensitivity analysis information discussed earlier. It will then be easier for decision makers to select, among a finite number of options their most preferred targets for inefficient units.

5. Conclusions

In this chapter, the use of prioritised target setting models was put forward as the basis for assessing performance targets for individual operating units. This family of models was originally introduced by Thanassoulis and Dyson (1988) and succeeds to a very large degree in encapsulating many of the target setting principles discussed in chapter three. These include the presence of managerial preferences over input/output improvements, the flexibility in treating the degree of controllability of the input/output factors and the ability to accommodate simultaneous improvements to the input/output factors.

A direct extension of the prioritised target setting model was put forward for assessing global organisational targets. This model which is provided in M4.4 seeks to generalise the industry production function models from the econometric literature. Despite the appeal of this model for setting global targets it is still insufficient for supporting resource allocation decisions. M4.4 gives little or no discretion to individual DMUs to "re-act" during the planning process. In this respect M4.4 is a "dictatoric" model as all the units of an organisation are aggregated, in isolation of their input/output mix, culture, technologies, etc., to a super-unit that represents an industry or organisation. This parallels the failed experiments for central planning experienced in the ex-communist countries (see Smith (1990)). More flexible and robust models for corporate/industry planning will be developed in chapter five.

The economic implications of the prioritised models developed were emphasised in the second part of this chapter. The assessment of prioritised targets encompasses the preferences of decision makers regarding the rate of improvement to inputs/outputs. It was argued, however, that these priorities also guide the projection of inefficient units onto facets of the efficient frontier that support different exchange rates between inputs/outputs. Simple sensitivity analysis in linear programming yields information on the association between the preferences on input/output improvements and the estimated exchange rates.

Despite their attractive features, the prioritised target models do not sufficiently address all aspects of target setting. Indeed, the assessment of performance targets using the aforementioned models would result in an improved use of target setting primarily as a *control* instrument in MUOs. The *planning* dimension of target setting is yet to be addressed in full. This is pursued in more detail in chapters five and six of the thesis.

- END OF CHAPTER FOUR -

Chapter 5

Centralised target-based planning models in multi-unit & multi-level organisations

1. Introduction

Chapter 5 builds on ideas developed in chapters three and four concerning the use of target setting for managerial control and planning in not-for-profit organisations. These earlier chapters pursued the development of a set of principles for target setting in MUOs and then explored models for setting targets consistent with these principles. Although the models explored, M4.2 and M4.3, accord in large measure with the target setting principles of chapter three they do not suffice for resource allocation purposes. The latter will be the main scope of this chapter.

The development of target setting processes is undoubtedly a crucial factor in the design of planning models. Planning models include aspects of resource allocation and decision support which need to be linked with the assessment of performance targets. An attempt towards this direction will be made in this chapter which is organised as follows.

The first section reviews existing methods of planning in multi-unit and multi-level organisations. Evidence from the existing literature shows that the current methods have not attracted widespread application in real life multi-level planning problems. In the second part of this chapter a target-based planning framework is proposed that seeks to overcome some of the problems identified previously. A goal programming model is developed for linking target setting and resource allocation processes adopting a centralised mechanism for co-ordinating decision making in MUOs. This implies that the representation of individual units and the co-ordination of the planning process are controlled by central management. The chapter concludes emphasising the centralised target-based planning model's ability to accommodate the principles of target setting.

2. Multi-unit & multi-level planning systems

The usefulness of decision support systems for resource planning is accepted within organisations. Planning models are used to support the deployment of resources, personnel, equipment and new technology in the best possible way in order to meet the organisational objectives at the micro (unit) and macro (organisational) level. We shall distinguish the so-called *centralised* and *decentralised* decision making processes as two extreme cases of planning in MUOs. It is argued that the choice of planning method relates mostly to the nature and power structure within organisations and, therefore, centralisation and decentralisation cannot be seen as alternative solutions to the same problem.

For example, it is difficult, if not impossible to apply decentralised planning mechanisms to military organisations because of their very rigid hierarchical structure. It is easier and more effective, however, for an economic and political organisation like the European Union to apply decentralised planning mechanisms among its member states. Bolton and Farrell (1990), developed a framework for comparing the performance and suitability of centralised and decentralised planning methods. They conclude that, despite the traditional economic belief that favours decentralised planning, the performance of alternative planning mechanisms is mainly affected by the nature of the planning problem concerned and not only by the selection of planning tools.

The choice of planning mechanisms in public MUOs is influenced by political factors, as illustrated by the case of the provision of local services by local authorities. Governmental policy seeks to influence local authorities to provide services in line with its economic and political desires. Local authorities, on the other hand, seek to implement policies according to their own priorities. The latter can create tension and conflict between local authorities and government, as well as among local authorities. Central government exercises its discretion on the allocation of central funds, as well as on legislative matters whilst, on the other hand, local authorities pressurise government by judicious choice of services to cut-back (e.g. free meals at schools) in order to avoid overspending or by imposing excessive levels of local taxation.

The case of planning in MUOs is an old one. There is an extensive management science literature related to multi-level planning problems, which is mostly theoretical with a limited number of real life applications such as, Ruefli (1971).

The idea put forward in these studies is that the design of large planning systems can be made by decomposing them into a number of smaller subsystems each with its own goals and policies. These models result in multi-level hierarchical systems where each level has a different degree of autonomy and authority over the organisational goals and objectives. As Anandalingam (1988) argues the problem in a multi-level system is how to ensure that all decision makers, acting according to their own goals, will achieve the overall system's goals.

Nijkamp and Rietveld, (1980) describe three principal problems of policy making in multi-level environments:

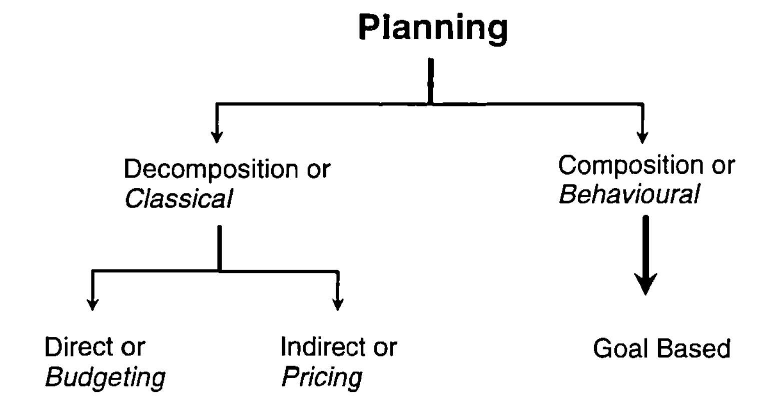
- *Interdependencies between the components of the system,*
- *Conflicts between various priorities, objectives and targets* **within** *individual components of the system,*
- *Conflicts between the priorities, objectives and targets* **between** *the various components of the system.*

Based on these three characteristics Nijkamp and Rietveld (1980), advocated the usefulness of multi-objective programming methodologies for addressing planning problems in these organisations. Presence of conflicting objectives at the global level requires use of multi-objective programming for achieving *compromise* solutions whilst conflicts between the organisational levels of the system require the presence of *co-ordinating* mechanisms throughout the organisational hierarchy. In principle, co-ordinating mechanisms consist of a common set of values and objectives which are agreed upon within organisation and it is usually the responsibility of central management (principal) to activate these mechanisms in order to monitor the planning process.

Each level of decision making has limited information on the objectives and policies adopted by other organisational levels or even among units within the same level (e.g. schools within the same geographical region). Acknowledgement of the *inter* and *intra* differences within MULOs implies that the system of "values" adopted by individual decision making entities is not always the same. It is difficult, therefore, to describe input-output relationships of different activity centres using the same "pricing" policy which in itself increases the role and significance of the co-ordinating mechanisms in the interior of MULOs. The exact form and characteristics of these co-ordinating mechanisms is open to different interpretations.

Two alternative schools of thought have emerged in the management science literature as solution methods of planning problems of MULOs. Ruefli (1974) calls them the *classical* and *behavioural* models whilst Sweeney *et al.* (1978) call them *decomposition* and *composition* approaches respectively. Figure 5.1 provides a pictorial representation of these methods.

Figure 5.1
Planning methods in MULOs



The classical or decomposition approach draws from the economic literature and is attached to the notions of optimality and ideal organisational problem. Organisations are assumed to have strictly pyramidal hierarchy with a global objective and various subunits. On the other hand the behavioural or composition approach assumes that strategy is driven by the subunits and not the ideal organisational problem. The behavioural model is based on a goal setting framework which seeks to obtain the most satisfactory solution for an organisation given its structure. This satisfactory solution will not always be an optimal solution in terms of the criteria used in the *classical* approach. Ruefli (1971) as an advocate of the behavioural school of thought states, that a firm with organisational structure, A and planning solution S(A), will need to obtain a planning solution S(B) if its organisational structure is to become B. Use of the classical approach of planning, however, would assume that the two planning solutions will always be the same, S(A)=S(B).

2.1. Decomposition approach for planning

Depending on the role and characteristics of the central authority as a co-ordinating body in the classical or decomposition approach two subcategories of methods can be identified. These are the *direct* and the *indirect* approaches also known as budgeting and pricing processes respectively by Burton and Obel (1977). Let us consider the case of budgetary planning for a set of schools controlled by a local education authority.

Using the *direct* approach for planning the central budget will be distributed directly among individual schools in monetary and non-monetary terms. Each school will then solve its own planning problem (provision of education services) and report opportunity cost (shadow price[1]) for the central budget. Having obtained these prices the central authority revises the initial distribution of resources to increase the rate of resource utilisation across all schools. When all the shadow prices of the schools are equal, the process has reached an optimal allocation. Examples of direct planning processes can be found in Ten Kate (1972) and Johansen (1978).

The *indirect* approach is based on a price based process for allocating resources. Distribution of resources takes place only after the optimal prices of resources have been obtained. The central authority starts off with the generation of provisional "prices" for the provision of services. Based on these prices each school estimates the optimal amount of budget required. The original prices are revised by the central authority until the schools acquire sufficient resources to support their activities. The mathematical rationale of this method is based on the original formulation of the decomposition algorithms in linear programming made by Dantzing and Wolfe (1961) and Baumol and Fabian (1964).

A more general comparison between the *resource* and *price* based approaches to planning can be found in Kornai's (1973) work where a parallelism is made between the capitalistic approach to planning (price based) and the centralised planned economies (resource based). This debate has been enhanced further by Land *et al.* (1993) where it is proposed that the

---

[1] The notion of shadow prices originates from the mathematical programming literature. The shadow price reveals the rate of change in the objective function for a unit change of the resources available to individual units.

main difference between capitalistic and socialistic planning are the different ways of response to change.

Ruefli (1974) provides a critical survey concerning the use of resource allocation models in MUOs alongside with other authors such as Sweeney *et al.* (1978). Some of these points are discussed next.

- *Structural problems*

As Freeland and Moore (1977) argue the direct or *resource directive* approach is not without its structural problems. The authors provide a series of theorems that show that the application of the *direct* approach is affected by the presence of multiple optimal sets of shadow prices for each subdivision during the solution process. This "pathological" phenomenon can seriously affect the solution process of the planning problem.

- *Development of effective pricing systems*

The estimation of "prices" for the inputs and outputs used in the planning process is another area of concern. In both decomposition (classical) methods of planning the use of prices is an essential part of the solution process. Burton and Obel (1989) argue that the prices attached to resources to initiate the solution process are arbitrary. Lewin and Morey (1981) also argue that all pricing systems in not-for-profit organisations suffer from the absence of a reference basis (e.g. market) for their estimation.

- *Representation of organisational objectives*

The current literature of multi-level planning also fails to make specific reference to the fundamental criteria of resource management notably equity, efficiency and effectiveness as discussed in chapter three. It is assumed that these criteria are represented through the interactive procedures between the central authority and the divisions during the allocation process. There is an implicit assumption of rationality in the behaviour, as well as the choice expressed by the individual "players" of the planning process. However, there is no specific reference as to how efficiency, effectiveness and equity will be represented in the planning process.

- *Limited applicability*

Ruefli (1974) argues, that the limited application of multi-level resource allocation methods to real life problems reflects, perhaps, their limited usefulness. It must also be noted that the only significant real life application of these methods by Manne (1973) on the Mexican economy has shown that the current methods have very abstract characteristics which limit their applicability.

2.2. Composition approach for planning

Evidently, the composition approach has been given much less attention by the research community and there is limited literature relating to this methodology. Sweeney *et al.* (1978) employed direct comparisons between the composition and decomposition methods whilst Ruefli is undoubtedly the most influential advocate of the composition approach. The goal programming based formulations representing the composition methods are attractive in the sense that they include organisational objectives, in the form of goals, reflecting objectives of the subunits and the global organisation. *The numerical estimation of these goals at the unit and/or global organisational level is, however, yet to be resolved. Moreover, the technological coefficients representing resource requirements per unit of output produced are not estimated in a well-defined and systematic way.* These are two areas that will be investigated further in the remainder of this chapter.

3. Rethinking multi-level planning

Multi-level organisations are organised into independent subsystems based on geographical and/or functional segments. For example, a country's education system has geographical subsystems, e.g. local educational authorities, and also functional subsystems, e.g. primary, secondary education. It is proposed, that in a complex system of this kind one needs to consider issues related with *planning* and *control* from the global educational, to the regional (Local education authority) and finally to the individual school level.

Alternative forms of centralised decision making seem to be the most common practice of planning in MULOs. Central authorities, often central government, are responsible for co-ordinating the planning decisions affecting individual divisions and ultimately the operating units. Decentralised planning on the other hand, is an issue with strong political, as well as managerial implications. There is no unique economic and managerial definition of decentralised planning. For some authors, the presence of a multi-level organisation is sufficient to consider the decision making system as decentralised. **In this research decentralisation is measured by the extent to which central and DMUs' management communicate (interact) in making decisions, as well as by the degree of autonomy**

that individual units have in planning their operations. One way of exercising this autonomy is to consider the freedom that individual units have in choosing their own technology (input/output mix) in delivering their services.

3.1. Equity, efficiency and effectiveness as objectives of planning

As it was argued earlier, the question of planning in MUOs concerns the best deployment of all types of resources in order to meet the organisational objectives. It will be erroneous, therefore, to proceed towards developing planning models without discussing the objectives that should guide the deployment of resources. The organisational mission is, undoubtedly, the starting point for decision making in an organisation. In Chapter one it was argued that organisations seek to operationalise their mission by stating organisational objectives. These objectives are quantified using targets whilst their achievement is supported by resource allocation and decision making. In a not-for-profit MUO environment the notions of *Effectiveness, Equity, and Efficiency* can be used as means for accomplishing the organisational mission.

Within this conceptual framework therefore, *Effectiveness, Equity, and Efficiency* can be considered as objectives that need to be maximised in order to support the organisational mission. Let us consider the provision of health services in a country. The organisational mission can be broadly defined as seeking to support and protect the well-being of citizens by improving their health standards. This mission can be fulfilled by deploying resources (e.g. doctors and nurses) within the whole country. The success of the resource deployment is to be determined by the extent to which the resources are used *effectively, equitably and efficiently.* The maximisation of these three components determines the extent to which the mission of the health system is fulfilled. It is noteworthy, however, that the definition/measurement of these concepts could vary between sectors (e.g. education/health) or between countries (e.g. UK and Greece).

In a seminal article, Savas (1978) sought to initiate a debate concerning the objectives in the provision of public services. Savas (1978) discusses the importance of equity as a resource allocation criterion, whilst arguing that management science has neglected in the past this important dimension when allocating resources in the public services. Sava's argument, however, draws upon practices experienced in Northern America. The European experience, in general, and in the UK, in particular, gives many examples of resource allocation based, perhaps solely, on equity criteria. An illustrative example of equity based studies is the large scale study by the by the Resource Allocation Working Party[2] (RAWP) in 1976 in the UK to develop a *relative need* based system for allocation of governmental grants for the provision of health. Similar examples, can be found in the allocation of resources in Education, Taylor (1989) and in Local authorities, Smith (1987). More recently, March and Schilling (1994) emphasise the presence of alternative methods for measuring equity in resource allocation problems in the public sector.

Considerable theoretical and empirical discussion can be found in the literature of studies focusing on the representation of the previous three objectives in resource allocation problems. Heiner *et al.* (1981), for example, sought to develop a decision support system for allocating resources for the care of mentally retarded people, in the USA. The authors made specific reference to equity, efficiency and effectiveness as objectives of resource allocation. However, their model building approach was based on the use of external information concerning the estimation of performance goals and also the efficient costing of the services provided. For example, an *a-priori* knowledge was assumed concerning the association between consulting time (costs) and patients' health improvement at all health units. Mandell (1991), suggested bi-level programming formulations for compromising on *equity* and *effectiveness* tradeoffs in allocating resources among the branches of a public library. Despite the advanced attempts launched by Mandell to measure equity using inequality indices his model suffered from the *a-priori* assumption (production function) of an association between population and library books' circulation for each library.

In summary, it can be argued that the use of effectiveness, equity and efficiency as planning objectives is limited, due to the restrictive assumptions that are used for their representation within the planning models. Furthermore, at the practical level, there are signs of over-emphasising some of these criteria at the expense of some others. For example, the national health system in the UK has shifted progressively from an *equity* based to an *efficiency* based system of resource allocation, that is based on the internal markets and the service contracts. Planning decisions based solely on any of these criteria can be "dangerous".

---

[2] The system of resource allocation of health funds has recently (1994) been reorganised by a research team from the University of York in the UK, Carr-Hill *et al* (1994).

Equity as sole criterion does not guarantee the best possible use of resources whilst efficiency in itself can increase inequality in the provision of public services.

> *The argument put forward in this research, however, is that the real challenge is in developing resource allocation systems where* equity, efficiency *and* effectiveness *are taken simultaneously into account.*

The representation of the three aforementioned resource allocation objectives into resource allocation processes is not sufficient for the development of operational planning systems. Drawing upon the current multi-level planning literature, it can be argued that issues related to the *co-ordination* of the decision process between headquarters and subunits; *priority assessment* of the relative importance of each of the three criteria; *quantification* of the three objectives in the solution process are the key success factors for any attempt to develop effective target-based planning systems.

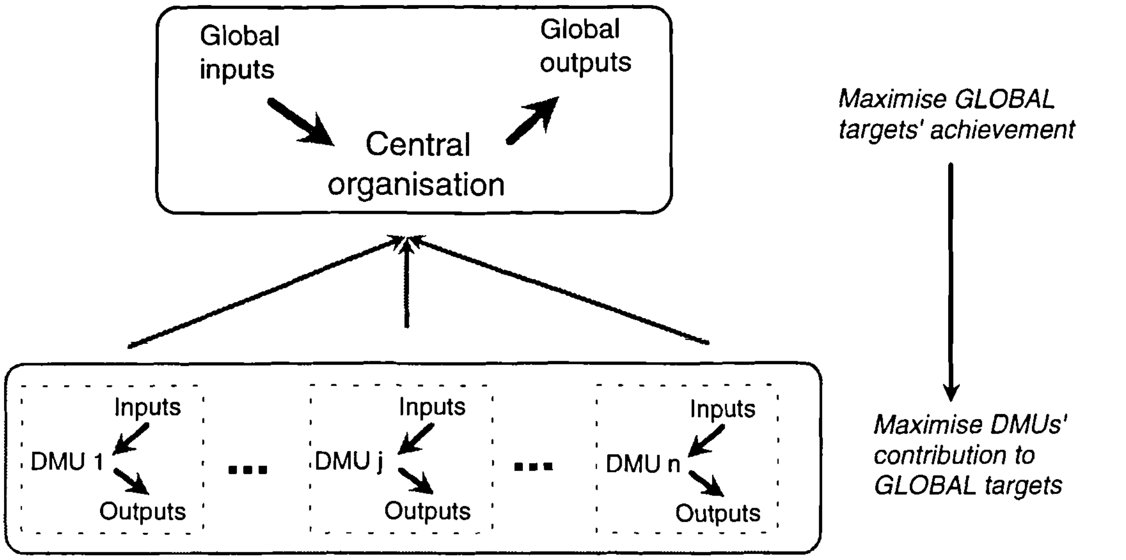
4. Centralised target-based planning

Centralised resource management is considered as the process where central management is responsible for the allocation and control of resources allocated to individual DMUs. Since these resources are obtained by central means (e.g. taxation) there is a need for public direction and accountability on their utilisation. Central co-ordination of the allocation process but more importantly auditing of the actual use of these resources by individual units is the typical route followed by central governments.

The target setting models developed in Chapter four provide a sufficient basis for investigating the extent to which individual DMUs utilise their resources by delivering outputs and services. The extension, however, of the auditing process towards resource allocation and decision support needs to consider all the DMUs of the organisation simultaneously. This would allow the planning process to take into account the global organisational targets, the global resource constraints and the internal communication between DMUs (resource re-allocation). A pictorial representation of this planning system within a MUO is given in Figure 5.2 and will be used as the basis for discussing the rationale of the centralised target-based planning system.

Figure 5.2

Rationale of the Centralised Planning system



The MUO represented in Figure 5.2 consists of a central co-ordinating mechanism that is responsible for controlling/allocating global resources to DMUs that operate similar but independent functions. Central management seeks to maximise the achievement of global input/output targets (assuming for the moment they are known). Individual DMUs are expected, therefore, to maximise their contribution towards the achievement of global organisational targets. This conceptual framework used to describe the operations of MUOs leads inevitably to questions concerning the operationality of the system, the assessment of global and DMU based targets and finally the management of interactions (resource re-allocation) between DMUs.

The models developed in Chapter four (M4.2 and M4.3) are sufficient when assessing targets for individual DMUs. The targets for the k[th] DMU using quantities of inputs x_i^k to deliver quantities of outputs y_r^k are given by the linear programming model M5.1.

$$\left.\begin{array}{ll} \underset{z_r^k,\theta_i^k}{Max} \quad \sum_{r\in O_c} P_r^k z_r^k - \sum_{i\in I_c} P_i^k \theta_i^k & \\ \sum_{j=1}^n \delta_j^k y_{rj} = z_r^k\, y_r^k & r\in O_c \\ \sum_{j=1}^n \delta_j^k x_{ij} = \theta_i^k x_i^k & i\in I_c \\ \sum_{j=1}^n \delta_j^k y_{rj} \geq \; y_r^k & r\in O_f \\ \sum_{j=1}^n \delta_j^k x_{ij} \leq \; x_i^k & i\in I_f \\ z_r^k \geq 1, \theta_i^k \leq 1, \delta_j^k \geq 0 & \end{array}\right\}(\forall\ k) \qquad \text{(M5.1)}$$

Where,

x_{ij}, y_{rj}	is the ith input and rth output of the jth DMU,
δ_j^k	is the factor contribution of the jth DMU to the targets of the kth DMU,
z_r^k, θ_i^k	are rates of improvement for the rth output and ith input of the kth DMU,
P_r^k, P_i^k	user specified preferences over the improvement of inputs/outputs of the kth DMU,
$I \equiv I_c \cup I_f$	is an index set representing inputs I=1,...,m,
$O \equiv O_c \cup O_f$	is an index set representing outputs O=1,...,s,
I_c, O_c	are subsets of the inputs/outputs to be improved,
I_f, O_f	are subsets of the inputs/outputs not to be improved.

The solution of M5.1 would yield targets $\left(\sum_j^n \delta_j^k x_{ij}, \sum_j^n \delta_j^k y_{rj}\right)$ for the inputs/outputs of each DMU k in isolation, which does not provide sufficient decision support for achieving the global targets of the organisation. A **simultaneous representation of all the DMUs** within the planning process of the MUO, as advocated in Figure 5.2, is necessary for capturing the global input/output character of planning. The representation has been achieved in the first set of constraints M5.2a of the planning model that is listed in M5.2 below. This representation is based on the formulation of M5.1 which has been converted into a goal programming form.

The activities of a multi-unit multi-level organisation can be aggregated and displayed by global levels of the inputs/outputs, which are allocated among or produced by individual operating units. Central management seeks to motivate individual units to maximise their performance insofar as the use of inputs and delivery of outputs is concerned. A way to motivate units in this direction is to set global performance targets for the organisation as a whole. The extent to which the organisation achieves these global targets is considered as a surrogate measure of its operational effectiveness. Individual units contribute to the global organisational targets which questions, however, the extent to which they contribute to the best of their ability. That is, the efficient contribution of individual units (see M5.1) to the global input/output targets can support the maximisation of the operational effectiveness of the organisation. The set of constraints in M5.2b can be used to represent the efficient contribution of individual units on the global input/output targets.

The planning process within a MUO is also subject to policy making type of constraints. As an example, a set of *balance of payment* constraints are proposed for linking the aggregate target achievements of commensurate inputs and outputs in the planning process. For example, balance constraints can be used for balancing the income-expense relationship in macroeconomic planning models of local government spending. These would allow central/local governments to operate under prespecified budget deficit/profit. The mathematical representation of these constraints is given in M5.2c, and it is similar with the balance constraints used previously in the industry targets model (see M4.4) of chapter four.

The foregoing discussion can be drawn together into the centralised target-based planning model M5.2. The model seeks to provide decision support that would maximise the achievements of global input/output targets by maximising the contribution of individual DMUs to those targets. Some notation is necessary to facilitate the mathematical formulation of the centralised target-based planning model.

- At the global organisational level a further distinction is made concerning the knowledge of the global input/output targets. It is anticipated that for a subset of controllable inputs I_V and outputs O_V management will be able to specify desired global targets whilst the global levels of the remaining inputs $\overline{I}_v$ and outputs $\overline{O}_v$ will be estimated through the solution of the model. It is expected that the distinction of the global controllability of inputs and outputs will apply to the inputs and outputs classified as controllable (I_C, O_C) at the individual unit level. Thus the inputs and outputs can be subdivided as follows: $I_v \cup \overline{I}_v \equiv I_c$ and $O_v \cup \overline{O}_v \equiv O_c$.
- Finally, the use of *balance constraints* to link the estimated global targets for the commensurate inputs and outputs will apply to a subset of commensurate inputs/outputs $I_B \subset I$ and $O_B \subset O$ respectively.

Using this notation the centralised target-based planning model is as follows.

Centralised target-based planning model (CTP) (M5.2)

$$\underset{\substack{p_i, p_r, n_i, n_r, d_i, d_r,\\ VX_i, VY_r, \delta_j}}{Min} \left\{ \sum_{j=1}^{n} \sum_{i \in I_c} \left(P_i^n \frac{n_i^j}{x_{ij}} + P_i^p \frac{p_i^j}{x_{ij}}\right) + \sum_{j=1}^{n} \sum_{r \in O_c} \left(P_r^n \frac{n_r^j}{y_{rj}} + P_r^p \frac{p_r^j}{y_{rj}}\right),\ \sum_{i \in I_v} P_i^g \frac{d_i^+}{GX_i} + \sum_{r \in O_v} P_r^g \frac{d_r^-}{GY_r} \right\}$$

Representation of individual DMUs

$$\left.\begin{array}{lll} \sum_{j=1}^{n} \delta_j^k y_{rj} - p_r^k + n_r^k & = y_r^k & r \in O_c \\ -\sum_{j=1}^{n} \delta_j^k x_{ij} + p_i^k - n_i^k & = -x_i^k & i \in I_c \\ \sum_{j=1}^{n} \delta_j^k y_{rj} & \geq y_r^k & r \in O_f \\ -\sum_{j=1}^{n} \delta_j^k x_{ij} & \geq -x_i^k & i \in I_f \end{array}\right\} (\forall \text{ unit k}) \qquad \text{(M5.2a)}$$

Effectiveness and global targets' achievement

$$\begin{array}{llll} -\sum_{j=1}^{n} \delta_j^1 x_{ij} - \cdots - \sum_{j=1}^{n} \delta_j^n x_{ij} + d_i^+ & = -GX_i, & (\forall i \in I_v) & \\ -\sum_{j=1}^{n} \delta_j^1 x_{ij} - \cdots - \sum_{j=1}^{n} \delta_j^n x_{ij} + VX_i & = 0, & (\forall i \in \bar{I}_v) & \text{(M5.2b)} \\ \sum_{j=1}^{n} \delta_j^1 y_{rj} + \cdots + \sum_{j=1}^{n} \delta_j^n y_{rj} + d_r^- & = GY_r, & (\forall r \in O_v) & \\ \sum_{j=1}^{n} \delta_j^1 y_{rj} + \cdots + \sum_{j=1}^{n} \delta_j^n y_{rj} - VY_r & = 0, & (\forall r \in \overline{O}_v) & \end{array}$$

Budget balance

$$\sum_{i \in I_B} \sum_{j=1}^{n} (\delta_j^1 + \cdots + \delta_j^n) x_{ij} - \sum_{r \in O_B} \sum_{j=1}^{n} (\delta_j^1 + \cdots + \delta_j^n) y_{rj} \leq B \qquad \text{(M5.2c)}$$

$$\forall i \in I_B \text{ and } \forall r \in O_B$$

$$\delta_j^k, n_i^j, n_r^j, p_i^j, p_r^j, d_i, d_r \geq 0,$$

$$VX_i \geq 0\ \forall i \in \bar{I}_v \text{ and } VY_r \geq 0\ \forall r \in \overline{O}_v.$$

Where,

n_i^j, p_i^j	are negative and positive deviation variables for the i^{th} input level of unit j,
n_r^j, p_r^j	are negative and positive deviation variables for the r^{th} output level of unit j,
d_i^+, d_r^-	are the positive and negative deviation variables from the global target of input $i \in I_v$ and output $r \in O_v$,
P_i^n, P_i^p	preferences over the minimisation of positive/negative goal deviations of i^{th} input,
P_r^n, P_r^p	preferences over the minimisation of positive/negative goal deviations of r^{th} output,
P_i^g, P_r^g	are the preference levels related to the global target of i^{th} input and r^{th} output,
x_{ij}, y_{rj}	are quantities of input i and output r of the DMU j,
GX_i, GY_r	are global target quantities of input i and output r imposed a-priori,
$VX_i, VY_r,$	are input $i \in \bar{I}_v$ and output $r \in \overline{O}_v$ global targets quantities to be estimated by the solution to the model,

B is a user specified constant concerning the balance between commensurable inputs and outputs in the planning model,

I_B, O_B are the subsets of commensurable inputs and outputs.

The model in M5.2 is a goal programming one. Goal programming has been previously suggested by Thanassoulis and Dyson (1992) in its purest form for estimating performance targets. However, their model aimed at assessing performance of individual units assuming some knowledge of "ideal" input/output targets. The formulation in M5.2 provides a framework with **no predetermined assumptions on individual units' achievements**. The structure of M5.2 is discussed in more detail next.

4.2. DMUs' representation in the CTP model

Each DMU in the model is represented by its own input/output constraints. For example, the constraint set M5.2a represents unit k. This constraint set is based on M5.1 regarding the comparisons made between the inputs/outputs of the assessed unit k $\left(x_i^k, y_r^k\right)$ and its composite unit $\left(\sum_j \delta_j^{k*} x_{ij}, \sum_j \delta_j^{k*} y_{rj}\right)$. The formulation in M5.2a differs, from the target setting model in M5.1 in the presence of the goal deviation variables for inputs n_i^k, p_i^k and outputs n_r^k, p_r^k of unit k. The allowance given to over and under achieving input/output "goals" in M5.2a has very important repercussions for the estimated targets of individual DMUs. As is discussed later, suitable formulations of the objective function of M5.2 can yield input augmentation and/or output reduction targets for supporting the achievement of the global organisational targets.

The representation of individual DMUs using separate sets of constraints M5.2 would also yield (in the optimal solution) efficient facets for projecting inefficient DMUs. The efficient facet of unit k, for example, will be identified from efficient units j with $\delta_j^{k*} > 0$. The criterion used for the selection of efficient facets, however, is an issue related to the nature of the objective function of M5.2. Solution to M5.2 selects facets, for inefficient DMUs, that would facilitate the achievement of global targets.

In ordinary DEA (see M5.1) each DMU is projected onto an efficient facet that would maximise its efficiency. In M5.2, in contrast, each unit is attached an efficient technology as chosen to facilitate the achievement of global organisational targets. This implies that inefficient DMUs may be projected on different efficient facets by the solutions of models M5.1 and M5.2 respectively.

4.3. Estimating global input-output targets

The global input/output targets are represented in M5.2 through the constraints in M5.2b. These constraints seek to aggregate the contribution of, say DMU 1, to the global targets of the ith controllable input ($\sum_{j=1}^{n} \delta_j^{1*} x_{ij}$) and of the rth controllable output ($\sum_{j=1}^{n} \delta_j^{1*} y_{rj}$). A further distinction is made between inputs/outputs with imposed global targets (GX_i, GY_r), and those with estimated aggregate targets, (VX_i, VY_r), by the solution to M5.2. Further elaboration needs to be made concerning the possible ways of estimating global input/output targets, (GX_i, GY_r), prior to the solution of M5.2.

Previous studies on multi-level planning models recognise the importance of this issue but assume prior knowledge of global target levels, Freeland and Baker (1975). A number of options are discussed below for assessing global organisational targets.

4.3.1. *Policy making managerial decisions*

Global organisational targets are often based on managerial judgment by central planning policy makers. As mentioned in chapter one, global resource levels influence the expectations on the global outcomes delivered by organisations. For example, governments set spending targets for various education functions derived from their annual budget. At the same time the government may set various output/outcome targets in education as a result of the committed resources (e.g. percentage of pupils going into higher education, increase in research quantity/quality in Higher Education).

A common situation that arises in public management concerns the incompatibility of the simultaneous achievements of cost reduction and outcome augmentation targets for individual DMUs. Furthermore, managerially and/or politically imposed targets can under or over estimate the capacity of an organisation to deliver services. Under-estimated global targets in M5.2b may result in resource underutilisation whilst over-estimated targets may result in unrealistic targets for individual units.

4.3.2. *Targets estimated by the global (industry) target model in M4.4*

One step towards systematising the assessment of global targets is to consult the best practices of the organisation under study, Farrell (1957). This would lead to using targets obtained by the industry target model in M4.4. As was shown in chapter four targets obtained by this model yield aggregate input/output levels that can be obtained based on the current input/output levels of observed units. Using the notation of chapter four a set of global targets can be established for each controllable input and output as described in M5.3.

Using the global target model developed in chapter four we can estimate upper and lower bounds for each input/output variable with imposed targets (GX_i and GY_r) in the formulation of M5.2. The maximum output and minimum input targets can be obtained by maximising (minimising) each output (input) variable included in M4.4. A number of linear problems equivalent to the total number of inputs/outputs need to be solved which will also yield lower bounds for outputs and upper bounds for inputs from the payoff table that will be created from the solutions to the sequence of optimisation problems. An algebraic representation of these constraints in given in M5.3.

$$\begin{aligned} LI_i \leq \; & GX_i \leq UI_i \quad \forall i \in I_v \\ LO_r \leq \; & GY_r \leq UO_r \quad \forall r \in O_v . \end{aligned} \tag{M5.3}$$

Where,

LI_i, UI_i are the lower and upper bounds for the global targets of the ith input,

LO_r, UO_r are the lower and upper bounds for the global targets of the rth output.

Alongside with methods employed for estimating these targets are the economic implications of the set of constraints that correspond to the global performance targets in the CTP model. These implications are explored next with reference to the optimal solution to model M5.2.

The shadow prices of the constraints in M5.2b yield information on the marginal costs and benefits of employing additional resources or requesting extra outcomes at the aggregate organisational level. This, however, applies to inputs/outputs where we have previous knowledge of desired targets. Opportunity costs/benefits of governmental spending in different sectors of public activity can be obtained. For example, let us consider governmental spending for health and education. One may desire to compare the opportunity cost of achieving 1% extra of a health global target (e.g. reduction in the length of waiting lists), with the opportunity cost of achieving a 1% extra of an education global target (e.g. reduction in the staff students ratio). Note here that the example does not intend to compare the nature of these targets, but the marginal cost (in £ or natural resources) of increasing their achievement.

The achievement of the global targets of inputs/outputs is also monitored by the balance constraint listed in M5.2c. This constraint seeks to balance the global targets of subsets of commensurate inputs and outputs such that the organisation can meet prespecified (B) targets of deficit or profit. These constraints are useful in planning problems with inputs and outputs of a financial nature (e.g. the central government's budgeting of the local authorities' finance).

4.4. Goal deviation variables & the objective function

The objective function of M5.2 contains goal deviation variables that correspond to the global input/output targets' achievement (d_i^-, d_r^+) and those that correspond to the inputs/outputs of individual DMUs $(n_{i,r}^j, p_{i,r}^j)$. To remove scale bias the under and over achievement variables have been standardised on a per unit basis of their corresponding input/output goal values. Elements of the objective function are also weighted by preferences that express central management's views on the relative importance of the input/output targets.

Various solution methods are available for linear goal programming problems, (see Ignizio (1983) and Romero (1991)). One possibility is to aggregate the goal deviation variables that correspond to both global and DMU goals. This would imply, however, that the preferences selected by the decision makers for global targets $P_{i,r}^g$ and for individual DMUs $P_{i,r}^{n,p}$ will encapsulate tradeoffs between the two sets of goals. Alternatively, a lexicographic solution approach can be adopted, where the solution to M5.2 will be obtained after a number of iterations, Ignizio (1982).

The deviations of the goals of the global targets levels are represented, in the objective function of M5.2 as typical goal programming variables $\underset{d_i, d_r}{Min} \sum_{i \in I_v} P_i^g \frac{d_i^+}{GX_i} + \sum_{r \in O_v} P_r^g \frac{d_r^-}{GY_r}$. It is noteworthy, however, that the deviation variables used for global inputs (outputs) represent an overachievement (underachievement) of the original input (output) global targets. An assumption is made, therefore, that individual DMUs would seek in principle to use more resources than available and to deliver less services than are desired by central management. This assumption can be relaxed by the simple modification of the goal deviation variables.

The part of the objective function that relates to the deviational variables of individual units $\underset{p_i, p_r, n_i, n_r}{Min} \sum_{j=1}^{n} \sum_{i \in I_c} (P_i^n \frac{n_i^j}{x_{ij}} + P_i^p \frac{p_i^j}{x_{ij}}) + \sum_{j=1}^{n} \sum_{r \in O_c} (P_r^n \frac{n_r^j}{y_{rj}} + P_r^p \frac{p_r^j}{y_{rj}})$ can be used for addressing the issue of resource transferability. Under and over achievement goal deviational variables, $n_{i,r}^j, p_{i,r}^j$ are used for the controllable inputs/outputs of each DMU. The presence of a two-way deviation variables implies that the problem can be solved by under or over achieving the observed input/output values for individual units.

This is a <u>fundamental</u> departure from DEA models[3] which assume that assessed targets should always contract in inputs and expand in outputs. In the context of performance measurement the goal deviation variables have stronger implications than in ordinary goal programming models. *This is because the "goal" levels in the right hand side of* M5.2 *are the observed input/output values of individual units. These are effectively "undesirable" goals in a performance measurement context and, therefore, the solution process should aid units to move away from their current input/output levels to more efficient ones.*

Alternative planning scenarios can be implemented concerning the targets' selection for individual DMUs using the sign and magnitude of the preferences $P_{i,r}^{n,p}$ in the objective function of M5.2. Three of these scenarios are listed in Table 5.1.

<u>Table 5.1</u>
Developing planning scenarios

Planning scenario	Preference policies
i. Output expansion & input contraction	$P_i^n, P_r^p \leq 0$ and $P_i^p, P_r^n \geq 0$
ii. Output & input expansion	$P_i^p, P_r^p \leq 0$ and $P_i^n, P_r^n \geq 0$
iii. Output & input contraction	$P_i^n, P_r^n \leq 0$ and $P_i^p, P_r^p \geq 0$

Each of the three scenarios listed in Table 5.1 have different managerial implications regarding the assessment of performance targets for individual DMUs. The first scenario covers the traditional case of target setting where inefficient units are expected to expand

---

[3] Thanassoulis and Dyson (1992) ideal targets' models is an exception but the user is expected to specify ideal targets for inputs/outputs. This need has been relaxed in the CTP model.

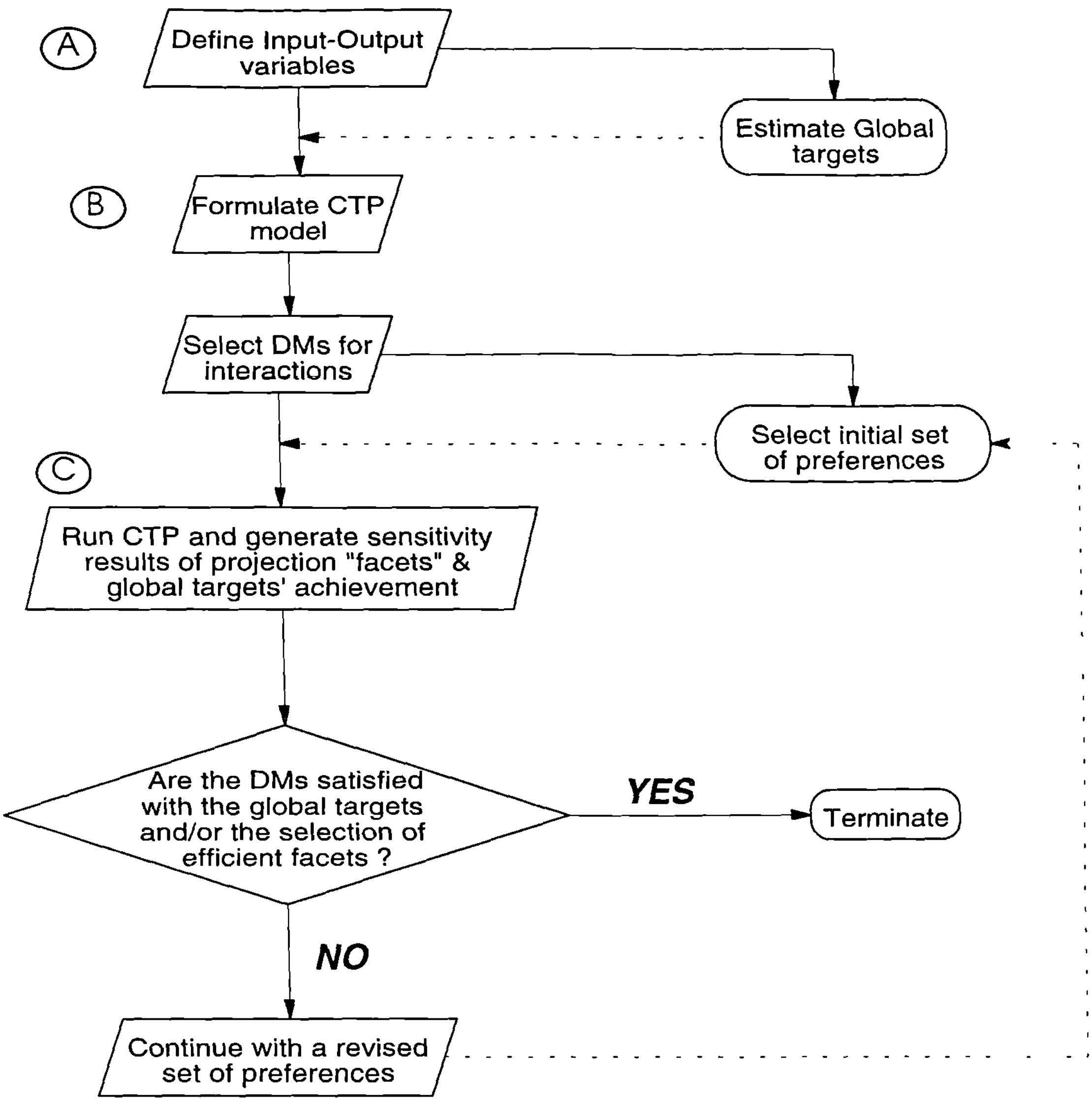
their outputs whilst reducing their inputs. This case corresponds to projections within area (1) of the production possibility set in Figure 3.2 in chapter three. The second scenario covers the case where extra resources are sought to be allocated to DMUs with the highest prospects for utilising these resources. Similarly this corresponds to the area (2) of the production possibility set of Figure 3.2. The last scenario concerns the case of simultaneous input/output reductions which corresponds to the area (4) in Figure 3.2. The strategy adopted for the projection of individual operating units can be guided by issues such as the economies of scale of the corresponding unit (DMUs operating under increasing returns to scale can be given extra resources anticipating higher rate of output returns).

The three scenarios discussed earlier can co-exist in the solution of the planning model. The latter implies that M5.2 is flexible enough to accommodate different strategies for individual operating units. As the solution of this planning model can take place after individual units have been assessed via ordinary DEA one can obtain different planning strategies for units with similar performance characteristics.

4.5. Operationalising the target-based planning model

The operational aspects of the mathematical development of the centralised target-based planning are discussed next. The solution of the CTP model requires various stages/phases that need to be linked. These stages are described in the algorithmic process below.

Centralised target-based planning (CTP)



A The solution process is initialised in **A** by the selection of input-output sets that describe the operating process at the unit and global organisational level. This would give the opportunity for estimating global input-output targets using some of the methods discussed earlier.

B The formulation of the centralised target-based planning model proceeds at this stage. An important element here is the selection of a panel of decision makers (DMs) that will facilitate the solution process.

C The interactive solution process constitutes the final part of the planning model. The main idea is that the DMs would be provided with information associated with the planning process and they will subsequently express their positive or negative impressions of the current solution. The information provided to decision makers can take alternative forms varying from the extent to which the global targets are achieved to the case of displaying the efficient peers for each inefficient unit in the analysis. Undoubtedly, the exact detail of this information will depend on the willingness of the decision makers to participate/understand the solution process.

The interactive scenario building process can be supported by the use of sensitivity analysis on the optimal solution of the CTP model. This can be addressed at two levels:

First, to investigate the allowable variation at the global level of targets, that are prespecified in the planning model (e.g. budget available for allocation), before new optimal solutions are obtained. Parametric analysis can also be used to explore the effects of simultaneous variations in the availability of more than one resource variable. For example, the allocation of resources among the school of a local education authority may investigate the implications of varying the total number of pupils within its geographical area, the abilities of the pupils admitted and finally the funds available for the operations of the schools.

Second, to express preferences over the minimisation of under/over achievement of goal deviation variables at individual DMUs. These preferences reflect penalties per unit of deviation from the right hand side goals and, therefore, the range of preferences that would leave the current optimal solution unchanged can obtained. In the case of schools this investigation would give better insights on the tradeoffs between the maximisation of pupils' achievements in different subjects, the resources allocated to the schools (e.g. pupils per student) and the abilities of the pupils admitted per school. A demonstration on how the preference levels in the objective function of DEA type of models can be used to explore alternative target levels for inefficient units has been given earlier on in chapter four (see section 4.3).

Model M5.2 is the first of a series of goal programming models that will be developed in this thesis to facilitate target-based planning mechanisms. The solution process adopted for goal programming models is a point of major academic controversies Hannan (1983), Ignizio (1982), Steuer (1983), and Min and Storbeck (1991). As Charnes and Cooper (1977) argue, however, the solution process selected for goal programming problems is *context* and *problem* dependent and therefore, no easy generalisations can be made. The most important implication, however, concerns the possibility of deriving *dominated* solutions either because the goals set at the first place were not ambitious or because the solution process adopted led to dominated solution. A typical remedy for this type of problem is the use of a two-phase solution process similar to the one adopted by Thanassoulis and Dyson (1992) in order to obtain *non-dominated* solutions. A more detailed discussion and remedial models concerning the presence of dominated solutions in goal programming problems are provided in chapter six.

4.6. Centralised target-based planning & the principles of target setting

The centralised target-based planning model was developed in an attempt to address problems of planning in MULOs from a performance measurement perspective. DEA models do not suffice for addressing planning problems as they concentrate solely on the performance of individual units without being able to encapsulate the interactions between individual DMUs. For example, the allocation of students of a given ability to a school needs to take into account the intake of other schools in the area such that a balance can be obtained. Similarly, the allocation of resources among the schools of the local education authority need to take into account the global resource availability within its annual budget. Clearly, the strength of DEA models to concentrate on a specific DMU at a time proves to be a weakness when planning problems are considered.

On the grounds of the DEA insufficiency to address these type of problems the centralised target-based planning CTP model was developed. This enhancement was based on the conceptual framework of effective target setting, which was proposed in chapter three (see section 3.2). This type of framework is necessary in order to keep systematic account of the operating characteristics of the proposed model. The mathematical formulation in M5.2 was put forward aiming to encapsulate the principles of target setting proposed in chapter three. In this section we assess the effect of these principles on the CTP model and the degree to which their representation was successful.

The CTP is a target setting and resource allocation model that gives particular emphasis on the achievement of global organisational targets. Individual DMUs contribute to the achievement of input/output global performance targets in line with the priorities of central management. The incorporation of *global performance targets* and also the use of *decision* *makers' preferences* are two important features of the CTP model. These two issues have been proposed as two of the principles of target setting in chapter three.

Each phase of the solution process of the CTP requires involvement of the decision makers which are expected to evaluate the desirability of the solutions obtained at each phase and also to express priorities for the phase to follow. At the operational level the decision makers' preferences will be considered at three core stages of the planning process:

i. In the development of an input-output model that represents the operations of the organisation.

ii. In the selection of aggregate target levels for inputs/outputs.

iii. In the selection of preferences over the improvement of inputs/outputs.

The selection of preferences for inputs/outputs has an interactive nature as decision makers may not have finalised ideas prior to the solution process. The iterative and interactive process for the selection of sets of preferences increases the likelihood for obtaining *managerially feasible* targets for individual DMUs. The latter is also a principle of target setting which gets particular importance in the solution process of the CTP. An effort must be made to ensure that central and local management share similar views insofar as the feasibility of the assessed targets are concerned.

The principle of value revelation, discussed previously in chapter three (section 3.2), was used to emphasise the need for incorporating organisational objectives in the assessment of performance targets. In a not-for-profit environment these objectives can be generalised using the concepts of *Efficiency, Effectiveness* and *Equity.* The representation of these objectives is compounded within the formulation of M5.2 due to its highly centralised character. It can be argued, however, that the achievement of the global organisational targets reflects the *degree of the operational effectiveness* of an organisation. The assessment of targets for individual DMUs shows the *efficient contribution* of each DMU to the achievement of organisational targets as it is represented by the set of constraints in M5.2a.

Finally, elements of *decision support* for allocating resources are incorporated in the formulation of the CTP model. The principle of *transferability* is supported by the set of constraints in M5.2 which allow the inputs/outputs to be re-allocated among individual DMUs. Note here that according to Mandell (1991) this principle is a basic component of *equity.* The predominate role of central management in determining the preference structure used in the planning process limits the ability of M5.2 to incorporate the full dimension of equity as a resource allocation objective. This is explored in chapter six with the decentralised planning model where the political and economic dimensions of equity are appreciated in full.

The reference made to the representation of the principles of target setting within the CTP model is not sufficient to guarantee their successful implementation in a real life problem. The application of CTP for allocating central grants to the Greek local authorities by Athanassopoulos (1994a) has revealed a number of critical factors that can affect the implementation of the planning process. Particular attention must be given to the selection and co-ordination of the decision maker's team that would contribute to the solution process. It is also important that the representation of the principles of target setting does not confronts the resource allocation and planning methods used by the organisation concerned in the past. Direct confrontation within a highly political environment (like the local authorities) may put at risk the viability of the proposed model and its solutions as it risks the stability of the organisation concerned. Resource allocation systems' reforms need to be used within a general process of reorganisation in order to eliviate the direct comparisons of those that benefit and those that loose at the end of the planning process. The gradual implementation of reallocation policies over extended periods of time should also be considered as a feasible option for supporting the implementation of the resource allocation system.

5. Conclusions

The motivation of this chapter rests on planning problems encountered in multi-unit and multi-level non-profit organisations. Throughout the chapter emphasis was placed on addressing two critical issues.

The **first** follows naturally from chapters three and four and concerns the enhancement of target setting mechanisms for linking *control* and *planning* processes. The models developed in chapter four for setting targets at the DMU and the global industry level are not sufficient for resource allocation. Therefore the planning characteristics of target setting were yet to be addressed operationally.

The **second** issue emanates from the limited support given by the current multi-level planning literature to the formulation of planning models that would encapsulate the principles of effective target setting.

The centralised target-based planning model presented here is a goal programming one. Similar methodology was followed by the so-called *composition* approach for planning developed by Ruefli (1971) and Freeland and Baker (1975). However, the CTP approach has a performance measurement and target setting orientation. The current version of the planning model presented in this chapter adopts a *centralised philosophy* of planning in MUOs. Targets, preferences, and technology are monitored by the central management without much flexibility for interacting with the DMUs included in the analysis.

The overall impression from the centralised target-based planning model is that it provides a useful framework for addressing planning problems in MUOs. The applicability of this type of planning model is associated with the existence of organisational environments with highly centralised structure and decision making processes. The "bottom-line" of not-for-profit organisations is highly centralised with governments or governmental bodies having discretion not only on the budgeting process but also on the rules of the game regarding the development/implementation of organisational strategies.

An alternative attempt for modelling target-based planning problems in MULOs will be made in chapter six. This attempt seeks to give a more active role to individual operating units during the resource planning process of the organisation and thus it characterises the decentralised target-based planning model, of DEA whilst retaining a central resource allocation role.

- END OF CHAPTER FIVE -

Chapter 6

Decentralised target-based planning in multi-unit & multi-level organisations

1. Introduction

Chapter 6 is concerned with the assessment of targets and the allocation of resources in *not-for-profit* multi-unit and multi-level organisations. Hitherto, the principles of target setting in this type of organisations have been discussed (chapter three) and prioritised target models at the DMU and global organisational level have been proposed (chapter four). The first attempt to link operationally the target setting with the resource planning process in MULOs was made in chapter 5 using the centralised target-based planning models. These models adopted a hierarchical and centralised approach for planning, where target setting and resource allocation decisions were co-ordinated by the central management.

The adoption of mathematical models for planning rests on the particular behavioural assumptions made about the structure of the organisations concerned. The models in chapter 5, for example, are based on *principal-agent* paradigms where the central management is the *principal* of the planning process with the individual DMUs being the *agents* that are expected to support the achievement of the global organisational objectives.

An alternative behavioural system of decision making is adopted in this chapter seeking to carry forward the debate between *centralisation* and *decentralisation* as alternative behavioural systems for estimating targets and allocating resources. That is, to maintain the target-based focus of the planning models developed previously whilst developing a new planning framework which gives more decision making discretion to individual DMUs during the planning process. This new framework, namely *decentralised* target-based planning (DTP) will provide an operational framework which:

- makes explicit reference to the resource planning objectives of individual DMUs and the global organisation,
- appreciates the impacts of tradeoffs between different resource planning objectives in the planning process,
- respects and facilitates the presence of diverse interests/preferences between the headquarters and the individual DMUs.

To pursue the decentralised planning framework this chapter is organised in the following manner: A network representation of MULOs is given in order to exhibit the pattern of communication, resource transferability, and administration in these organisations. Using some of the fundamentals of network theory, a model building process, will be implemented in order to represent the resource management objectives within the network formulation. This formulation will lead to the development of a goal programming model. A subsequent discussion on the formulation of the decentralised target-based planning model will focus on both the organisational and operational aspects of the model.

2. Decentralised target based planning

The evolution of decentralised decision making in MULO is a natural by-product of the debate among political scientists and organisational theorists which concern the structure of decision making in *not-for-profit* organisations, Bennett (1980). Numerous theoretical and empirical studies from political science, Brooke (1984); organisational theory, Govindarajan (1988); economics, De Groot (1988) and Bolton and Farrell (1990) and operational research, Burton and Obel (1988) and Van de Panne (1991), can be found *inter alia* debating the differences between *centralised* and *decentralised* decision making and planning. The evidence provided in these studies shows no clear answer as to whether *centralised* is more advantageous than *decentralised* planning or *vice versa.* The thesis does not intend to provide an exhaustive comparison between the two systems of planning as its main emphasis is to provide models to support either type of planning.

The target-based planning model developed in chapter 5 sought to support *centralised* planning processes. Its rationale is that central management has the primary role in allocating resources and targets among individual units. Despite the elegant features of the CTP model, notably resource transferability, value revelation, decision makers' preferences and incorporation of organisational objectives, organisations are viewed as highly hierarchical entities. For example, the "technology" (rates of conversion of inputs into outputs) that each DMU employs in order to support the achievement of the global organisational objectives is selected centrally. Moreover, the CTP model is structured to accommodate the decision making preferences on organisational objectives from a single level of decision making which limits the participation of DMUs' management in the planning process.

The objective in this chapter is to relax *inter alia* the top-down selection of "technology" as advocated in chapter 5 by letting individual units choose their "own" efficient technology prior to the resource allocation process. This new development requires the representation of DMUs within the planning process as independent entities that interact/compete for centrally allocated resources. The selection of efficient technologies for individual DMUs, can be obtained using the concepts of efficient rates of transformation as estimated from the prioritised target setting models in chapter 4.

2.1. A network representation of MULOs

There is a growing literature in organisational theory makes use of principles of network theory as a vehicle for analysing issues relating to the operations of organisations, Axelsson and Easton (1992). Multi-unit and multi-level organisations provide an ideal environment for developing/applying principles of network theory. Typically, network theory is used to represent information flow and communication routes to support the design of effective information and decision support systems. In this research, however, the network representation focuses on the entire operation of an organisation facilitated by the use of input-output frameworks.

Let us assume that a set of activity centres J=1,..,j,..,n operate within a MULO. Each unit uses a vector $X \in \Re_+^m$ of input quantities to deliver a vector $Y \in \Re_+^s$ of output quantities where x_{ij}, y_{rj} is the observed level of the i^{th} input and r^{th} output of DMU j. A complementary notation will be introduced in this chapter with ϕ_{ij}, ψ_{rj} denoting unknown target quantities of the i^{th} input and r^{th} output level of the j^{th} DMU. **The estimation of these quantities is the main objective of the chapter.**

At the global organisational level organisational targets Gx_i, Gy_r can be defined for the input/output variables of the organisation as indicated in chapter 5. Let us assume that for each DMU, included in the network, the rates of transformation of input quantities into outputs and services are known. **Technological coefficients (β_{ij}, α_{rj}) for the i^th input and r^th output of DMU j can be defined that indicate that efficient production of one unit of output r would require β_{ij}/α_{rj} units of input i.**

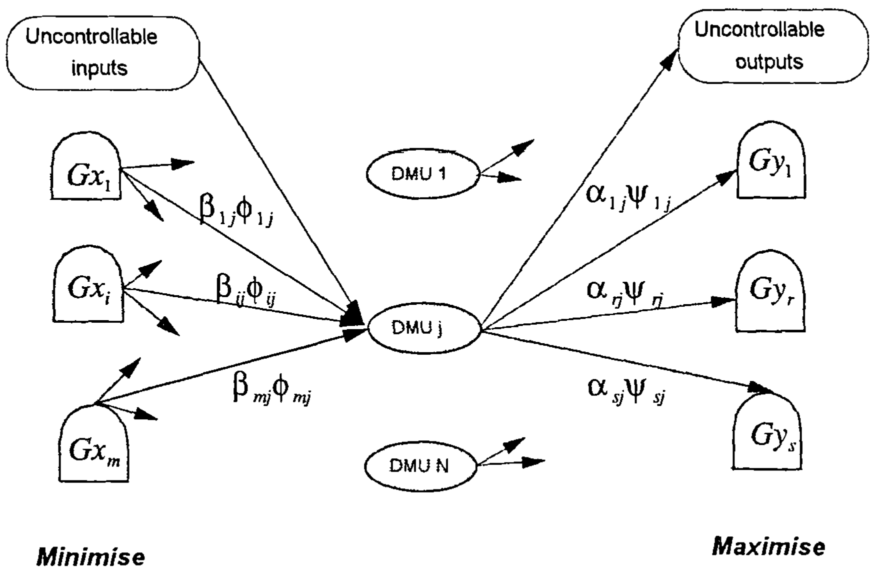
The estimation of these technological coefficients can be made using the prioritised target model (M4.4) developed in chapter 4. Model M4.4 is reproduced below for ease of reference. Let us denote the subsets (I_c,O_c) of index sets I=1,...,m and O=1,...,s of inputs and outputs sought to be improved and (I_f, O_f) the subset of inputs/outputs without improvement intentions. Let us also assume that user specified preferences, P_i^-, P_r^+, over the improvement of individual inputs/outputs are available. The model solved for assessing prioritised targets is M4.4 reproduced here as M6.1.

$$
\begin{array}{llll}
\underset{\alpha_r,\beta_i,\gamma_r,\zeta_i}{Min} & \sum_{i\in I_c}\beta_i x_{ij_o} + \sum_{i\in I_f}\zeta_i x_{ij_o} - \sum_{r\in O_c}\alpha_r y_{rj_o} - \sum_{r\in O_f}\gamma_r y_{rj_o} + \sum_{r\in O_c}P_r^+ - \sum_{i\in I_c}P_i^- & & \\
\textit{subject to} & & & \\
& \sum_{i\in I_c}\beta_i x_{ij} + \sum_{i\in I_f}\zeta_i x_{ij} - \sum_{r\in O_c}\alpha_r y_{rj} - \sum_{r\in O_f}\gamma_r y_{rj} \geq 0 & (\forall j=1,\ldots,n) & \text{(M6.1)} \\
& \alpha_r y_{rj_o} \geq P_r^+ & (\forall r\in O_c) & \\
& \beta_i x_{ij_o} \geq P_i^- & (\forall i\in I_c) & \\
& \alpha_r,\beta_i \text{ are free variables; } \gamma_r,\zeta_i \geq \epsilon. & &
\end{array}
$$

The model in, M6.1, needs to be solved separately for each unit j and, therefore, its formulation is an instance of the linear programme relating to DMU j_O. For future reference we shall denote, $\alpha_{rj}^*, \beta_{ij}^*, \gamma_{rj}^*, \zeta_{ij}^*$, as the optimal solution to M6.1 that corresponds to the instance of M6.1 relating to DMU j. The advantages of M6.1 have been discussed extensively in chapter four concerning the estimation of non-zero marginal rates of substitution for all inputs/outputs included in the analysis.

The operating process of the units of an organisation can be represented via a network formulation. The network will consider as inflows the input quantities used by individual DMUs and as outflows the output quantities produced. The conversion process of inputs into outputs will be enhanced using the efficient rates of transformation between inputs and outputs of each DMU, as obtained by M6.1. The network representation of MULOs is exhibited in Figure 6.1.

Figure 6.1
Input-Output network flow models



A network system has *sources, routes,* and *destinations* as its main components. In the context of this study these components are represented as follows:

- There is a set of input production components that operate as *sources.*
- A set of operating units constituting alternative routes of input flow and output generation.
- A set of outputs as final *destinations*; delivered by individual activity centres.

The organisation conglomerate represented in Figure 6.1 is assumed to be subject to resource allocation decisions controlled by a central authority (e.g. local education authority (LEA), district authorities, and department of environment for local authorities).

The left hand side (shaded) flows, ($\beta_{ij}\phi_{ij}, \forall i,j$), relate to the controllable inputs used by the organisation, where β_{ij} are the technological coefficients for input i of DMU j estimated by M6.1. The aggregate quantity, Gx_i, of each input needs to be allocated among the n DMUs. This allocation is, in principle, feasible for only controllable inputs/outputs which are represented separately in Figure 6.1. The arrow paths in the network represent the possible flow of resources. The unit costs for reallocating resources between the operating units is assumed to be zero. The right hand side flows, ($\alpha_{rj}\psi_{rj}, \forall r,j$), relate to the outputs delivered by the organisation. An assumption has been made that the inputs (resources) used in the network must be minimised whilst the outputs/outcomes are maximised. Undoubtedly, these assumptions can be modified in the context of specific applications (e.g. increase spending for the reduction of crime).

At the operational level, the monitoring of the network flows will be considered regarding the identification of the **best possible routes for allocating resources and delivering maximum output quantities**. The selection of alternative routes (DMUs) for allocating resources in order to produce outputs/services will be guided by the various planning objectives included in the system (efficiency, effectiveness and equity).

2.1.1. *DMUs' operation in a network structure*

The operation of individual units is represented within the network structure as input inflows and output outflows. Each DMU has its own inflow requirements in order to generate its corresponding outflow. The inflow-outflow process can be appraised by the law of ***flow conservation*** which balances the inputs received by each DMU with its outcomes generated. In a transportation network, for example, the flow conservation requires all incoming goods to equal outgoing goods.

Given a total inflow $\sum_{i=1}^{m} \beta_{ij}^{*}\phi_{ij}$ and outflow $\sum_{r=1}^{s} \alpha_{rj}^{*}\psi_{rj}$ of the jth DMU the mathematical representation of the flow conservation law is given in M6.2.

$$\sum_{i=1}^{m} \beta_{ij}^{*}\phi_{ij} - \sum_{r=1}^{s} \alpha_{rj}^{*}\psi_{rj} - t_j = 0 \qquad \forall j = 1,\ldots,n, \tag{M6.2}$$

where

$\beta_{ij}^{*}, \alpha_{rj}^{*}$	are technological coefficients of the ith input and rth output of the jth DMU,
ϕ_{ij}, ψ_{rj}	are unknown quantities of ith input and rth output of the jth DMU,
$\sum_{i=1}^{m} \beta_{ij}^{*}\phi_{ij}$	weighted total inflow of resources to DMU j,
$\sum_{r=1}^{s} \alpha_{rj}^{*}\psi_{rj}$	weighted total outflow of outputs from DMU j,
t_j	accounts the *waste* or inefficiency in the operation of jth unit.

Efficient operation of unit j implies zero value for t_j, while for inefficient DMUs, one can decompose this inefficiency into different terms (e.g. technical and scale) as was discussed in chapter 2. The purpose of the technological coefficients β_{ij} and α_{rj} which are obtained by M6.1 is threefold.

i./ To convert the incommensurate inputs/outputs into a commensurate scale,

ii./ to aggregate multiple inputs and multiple outputs into a commensurate total inflow and outflow quantity,

iii./ to introduce the element of *efficiency* as the ratio of technological coefficients, β_{ij}/α_{rj}, which represents the per unit of input i efficient requirement for producing one unit of output r.

The conservation flow equations in M6.2 emphasise the independent representation of individual DMUs within the decentralised planning process using their own set of rates of transformation of inputs into outputs. This formulation will be used as the main component of the DTP model as each DMU selects its own efficient rates of transformation prior to participating in the allocation of central grants.

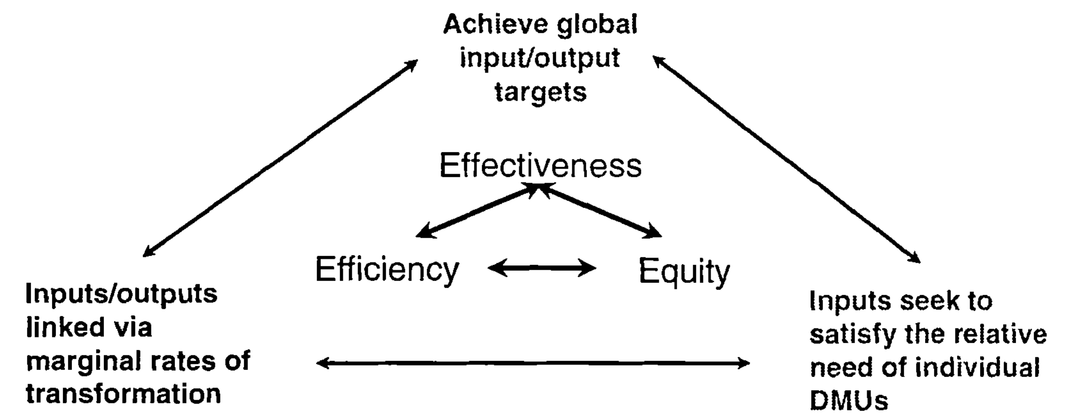
2.2. Objectives and co-ordination in MULO networks

The flow conservation equations convey a mathematical representation of the process of converting input quantities into products and services. However, the presence of the flow equations in M6.2 needs to be enhanced further in order to accommodate other objectives within the network flow management.

The network in Figure 6.1 provides a pictorial representation of how individual DMUs are associated within the global organisation in MULO planning problems. The co-ordination of different levels of decision making in the planning process needs to be considered in order to accommodate the potentially diverse interests of different levels of management in MUOs and, furthermore, to achieve explicit representation of the resource planning objectives in the modelling process.

Three fundamental resource management objectives, namely effectiveness, efficiency and equity, were previously discussed in the model building process of the centralised planning model in chapter 5. The same objectives are sought to be represented in the formulation of the decentralised planning model in this chapter. The additional aim here is to provide explicit representation of the three objectives of resource allocation within the planning model. A pictorial representation of these objectives is provided in Figure 6.2 below which will be used for discussing the rationale of the decentralised target-based planning model prior to its mathematical formulation in M6.4.

Figure 6.2
Objectives of resource management



The three objectives listed in Figure 6.2 seek to support *not-for-profit* organisations to accomplish their mission. The satisfaction of each of these objectives requires appropriate decisions and allocation of resources. A simultaneous maximisation of the effectiveness, efficiency and equity within a MUO may not be always feasible due to limited resources. Thus the planning process should aid management to find the most satisfactory levels of achievement of the three objectives. The rationale to include the three objectives within the planning formulation is discussed in more detail next so as to motivate the mathematical formulation of the decentralised planning model that follows in M6.4.

The <u>effectiveness</u> of a MUO can be quantified through the extent to which the organisation achieves its global targets. Issues related with the definition and quantification of global organisational targets have already been discussed in chapter 4 and 5. The overall amount of available resources and desired outcomes in a MUO can be used to construct a system of global target goals. These goals are incorporated in the decentralised target model using the set of constraints M6.4a.

The use of <u>efficiency</u>, as an objective for managing resources, promotes the idea that the achievement of global performance targets needs to be pursued by minimising the amount of wasted resources by each DMU. This issue has already been raised in the conservation flow equations (M6.2) in the network representation of MUOs in Figure 6.1. **Using the marginal rates of transformation between inputs/outputs, obtained from DEA in (M6.1), the DTP model (M6.4) will seek to find the most appropriate quantities of inputs and outputs to maximise the efficiency of individual DMUs.** This is the reverse process of what is typically followed in DEA studies where the input/output quantities of individual DMUs are known. The mathematical representation of efficiency within the DTP model is made using the set of constraints in M6.4b.

The importance of equity as an objective for allocating resources in not-for-profit organisations has already been discussed in chapters 1 and 3. Empirical studies on the distribution of public services, Wilson and Gibberd (1990), appreciated the difficulties for simultaneous representation of the three objectives in a resource planning process and proceeded to give exclusive priority to equity. The studies by Mandell (1991) and Heiner *et al.* (1981), (see section 3.1 of chapter 5), sought to encapsulate more than one objective in the allocation process utilising multi-objective programming methods. An enhanced attempt is made via the DTP model in M6.4 to consider equity along side with effectiveness and efficiency in the allocation of resources.

In the basic formulation of the DTP model in M6.4 equity will be represented estimating the *relative need* of individual units to obtain resources. One way of estimating relative need is to use multiattribute additive value (MAV) functions. Given an index set E of equity criteria with x_{ej} being the score of the j^{th} DMU on the e^{th} equity criterion we can define the relative need of the j^{th} DMU using the formulae in M6.3.

$$RN_j = \sum_{e \in E} w_e \frac{x_{ej}}{\sum_{j=1}^{n} x_{ej}} \qquad \text{(M6.3)}$$

where

w_e is a user specified weight factor representing the relative importance of e^{th} equity criterion in assessing equity.

The sociodemographic characteristics in the surrounding area of units providing public services are chiefly used as relative need criteria. For example, the allocation of governmental funds to geographical segments (e.g. district health authorities) can use as equity criteria, the population characteristics of each region, deprivation, mortality and morbidity indices. The estimation of the relative need of each region reflects the weighted proportion that each region contributes to the overall score of each particular equity variable (e.g population). The representation of equity in the DTP model is made using the set of constraints in M6.4c.

3. Goal programming for decentralised target-based planning

Having introduced the rationale of the decentralised planning model in the previous section the next step is to develop an operational form that would solve the DTP model. The two most important features that need to be considered in the DTP model are:

- *its multi-objective nature,*
- *its support for decentralised decision making behaviour.*

Central to the idea of the development of M6.4, is that individual units should be represented with their **own technology** in the resource allocation process. Moreover, the model should yield direct allocation of resources, as the **variables** of the problem are the controllable inputs (ϕ_{ij}) and outputs (ψ_{rj}) of individual operating units.

The multiple objective nature of the problem is addressed using goal programming. A specific goal is assigned to each objective and the model will seek to find the most preferred combination of inputs/outputs that would maximise the achievement of the goals.

The formulation of the goal programming model in M6.4 requires definition of some notation concerning the variables used in the model. Let us consider the index set of inputs I=1,...,m and outputs O=1,...,s with (I_c, O_c) being inputs/outputs that are sought to be improved and (I_f, O_f) inputs/outputs without expectations to improve. An additional index set E is also defined and concerns the factors used to estimate the relative need of DMUs for inputs as deduced from the equity objective. The nature of the criteria for defining equity is flexible and some of them can be the uncontrollable inputs of the index set I_f. On the other hand, however, there is no restriction that all the equity criteria should be used within the input-output model of efficiency and effectiveness.

Decentralised target-based planning (DTP) model (M6.4)

$$\underset{\substack{\psi_{rj},\phi_{ij},t_j,\\ D_i^{-,+},D_i^{-,+},\varepsilon_{ij}^{-,+}}}{Min} \left\{ \sum_{r\in O_c} \frac{P_r^+ D_r^+ + P_r^- D_r^-}{Gy_r} + \sum_{i\in I_c} \frac{P_i^+ D_i^+ + P_i^- D_i^-}{Gx_i},\ \sum_{j=1}^{n} P_j t_j,\ \sum_{i\in I_c}\sum_{j=1}^{n} \left(P_{ij}^{\varepsilon-}\varepsilon_{ij}^- + P_{ij}^{\varepsilon+}\varepsilon_{ij}^+\right) \right\}$$

Effectiveness	$\sum_{j=1}^{n} \psi_{rj} + D_r^- - D_r^+ = Gy_r$	$(\forall r \in O_c)$	**(M6.4a)**
	$\sum_{j=1}^{n} \phi_{ij} + D_i^- - D_i^+ = Gx_i$	$(\forall i \in I_c)$	
Efficiency	$\sum_{i\in I_c} \beta_{ij}^* \phi_{ij} - \sum_{r\in O_c} \alpha_{rj}^* \psi_{rj} - t_j = F_j^*$	$(\forall j)$	**(M6.4b)**
	$F_j^* = -\sum_{i\in I_f} \zeta_{ij}^* x_{ij} + \sum_{r\in O_f} \gamma_{rj}^* y_{rj}$		
Equity	$\phi_{ij} - \varepsilon_{ij}^+ + \varepsilon_{ij}^- = Gx_i * \left(\sum_{e\in E} w_e \frac{x_{ej}}{\sum_{j=1}^{n} x_{ej}} \right)$	$(\forall i \in I_c)$ $\forall j$	**(M6.4c)**
Policy constraints	$L\phi_{ij} \le \phi_{ij} \le U\phi_{ij}$	$(\forall i \in I_c)$	**(M6.4d)**
	$L\psi_{rj} \le \psi_{rj} \le U\psi_{rj}$	$(\forall r \in O_c)$	
	$\phi_{ij}, \psi_{rj}, D_i^{+,-}, D_r^{+,-}, t_j, \varepsilon_{ij}^{+,-} \ge 0.$		

Where,

ϕ_{ij}	amount of i^{th} input allocated to the j^{th} unit to be determined by the DTP model,
ψ_{rj}	amount of r^{th} output allocated to j^{th} unit to be determined by the DTP model,
x_{ij}, y_{rj}	observed uncontrollable input/output levels of j^{th} DMU, $(i \in I_f, r \in O_f)$,
Gx_i, Gy_r	specified global targets for the i^{th} input and r^{th} output,
D_r^+, D_r^-	over/under achievement goal deviation variables for r^{th} output global target,
D_i^+, D_i^-	over/under achievement goal deviation variables for i^{th} input global target,
$\alpha_{rj}^*, \beta_{ij}^*$	technological coefficients for controllable i^{th} input and r^{th} output, <u>always >0</u> estimated by M6.1,
$\gamma_{rj}^*, \zeta_{ij}^*$	technological coefficients for uncontrollable inputs and outputs estimated by M6.1,
t_j	reflects the inefficiency for the j^{th} unit,
F_j^*	is a fixed component attributed to uncontrollable inputs/outputs of j^{th} unit,
$\varepsilon_{ij}^+, \varepsilon_{ij}^-$	under and over achievement variables of the *equity* goals for the i input,
w_e	weight of the relative importance of e^{th} factor in the equity formulae,
$L\phi_{ij}, U\phi_{ij}$	user specified lower and upper bounds for the i^{th} input of the j^{th} unit,
$L\psi_{rj}, U\psi_{rj}$	user specified lower and upper bounds for the r^{th} output of the j^{th} unit,
$P_i^{-,+}, P_r^{-,+}$	user specified preferences of the i^{th} input and r^{th} output global target,
P_j	user specified preferences over the minimisation of inefficiency of the j^{th} unit,

$P_{ij}^{-\varepsilon}, P_{ij}^{+\varepsilon}$ user specified over the minimisation of the under and over achievement of *equity* targets in the allocation of i^{th} input of j^{th} unit.

Model M6.4 is a goal programming based one and consists of four sets of goal constraints that represent the multi-objective nature of the problem. The objective function of M6.4 is made of three groups of goal deviational variables. Each group reflects one particular dimension of resource management, namely effectiveness (M6.4a), efficiency (M6.4b) and equity (M6.4c) as discussed previously. The quantitative representation of the three objectives of resource management and the structure of M6.4 are next discussed in more detail.

3.1. Effectiveness within the decentralised planning model

The achievement of global organisational targets constitutes the main purpose of the planning model in M6.4. The concept of global organisational targets and their estimation has already been met in chapters 4 and 5. The planning scenario developed in this chapter seeks, however, to give a more decentralised form to the planning question and, therefore, it is important to clarify the assumptions made concerning the estimation of global targets in the decentralised target-based planning model.

Some further clarification is necessary concerning the relation between the global performance targets and organisational effectiveness. The difficulties to define and, more importantly, to measure organisational effectiveness have led some researchers to characterise any attempt for its quantitative representation as utopic. Such debate has been avoided in this research by making a distinction between the concept of *operational* and *ideal effectiveness* of an organisation. It is argued that operational effectiveness can be assessed quantitatively by using the global performance targets (Mandell (1991)), while the ideal effectiveness has a more qualitative character and is not pursued any further in this research.

Global targets are used to reflect the *operational effectiveness* of organisations. This is addressed by minimising the goal deviation variables included in the set of constraints in M6.4a. These constraints include inputs/outputs which are considered as controllable from the central management's point of view. The nature of the global performance targets may involve resource levels such as total number of people employed in the MUO network (e.g. number of doctors within the geographical segment of the network of DMUs) or the total spending target as is estimated by the treasury department at the beginning of the fiscal year. On the output side the operational effectiveness of the system comprises targets for service levels in a quantitative manner, such as the number of pupils enrolled within a network of schools, and also in a qualitative manner such as the abilities of pupils admitted and the value added achievements within a given geographical area. The systematic methods (global target setting using model M5.3) discussed in chapter 5 can also be consulted in order to obtain quantitative representation of the global targets.

3.2. Efficiency & individual unit representation

Individual units are represented in model M6.4 via the set of conservation flow equations M6.4b. In ordinary DEA these equations represent the difference between the weighted sum of inputs and outputs for each operating unit, with the weights being the variables of the problem. In the resource allocation problem, however, the weights ($\beta^*_{ij}, \alpha^*_{rj}, \gamma^*_{ij}, \zeta^*_{rj}$) are known whilst the inputs/outputs are the variables (ϕ_{ij}, ψ_{rj}) of the model.

The technological coefficients ($\beta^*_{ij}, \alpha^*_{rj}, \gamma^*_{ij}, \zeta^*_{rj}$) estimated in M6.1 reflect decision makers' preferences over input/output improvements. **In this case preferences seek to encapsulate the views of decision makers at the same level of administration as the operating units. Thus, the technological coefficients used in M6.4b represent an individual DMUs preferred directions of projection on the efficient frontier.** This is a very important feature of model M6.4 as it makes explicit reference to the diverse priorities of planning emanating from different decision making levels in MULOs.

The decentralised planning process accommodates efficiency using the conservation flow equations in M6.4b which are next discussed in more detail.

The first component of the equation $\sum_{i\in I_c} \beta^*_{ij}\phi_{ij} - \sum_{r\in O_c} \alpha^*_{rj}\psi_{rj}$ contains the variable levels of controllable inputs/outputs (ϕ_{ij}, ψ_{rj}) that will be determined by the solution of M6.4 in such a way that they will match the uncontrollable part of the equation $F^*_j = -\sum_{i\in I_f} \zeta^*_{ij}x_{ij} + \sum_{r\in O_f} \gamma^*_{rj}y_{rj}$ which is a constant term as all of its parts are known. For inputs and outputs that are considered as non-controllable (I_f, O_f), their reallocation between operating units is not feasible[1]. Their actual levels x_{ij}, y_{rj} and their technological coefficients ζ_{ij}, γ_{rj} are treated, therefore, as a fixed term, F_j, in the efficiency equation. Finally, the t_j term accounts for the amount of inefficiency allowed for each operating unit. The inefficiency term, t_j, will help to control the extent to which the individual DMUs will be allowed to operate inefficiently after the allocation of resources.

The potential presence of non-unique sets of weights, $(\beta_{ij}, \alpha_{rj}, \gamma_{rj}, \delta_{ij})$, for individual DMUs is a problem that needs to be discussed. The problem was mentioned, previously, in chapters 2 and 4 and concerns almost exclusively efficient DMUs. The allocation of resources, using M6.4, will depend on the set of efficient weights chosen for the representation of efficient DMUs in the efficiency objective in M6.4b.

The problem of multiple optimal sets of weights can be alleviated using our knowledge about the minimum and maximum optimal value that each weight factor can take in the case of multiple optimal solutions. Let us denote $\beta_{ij}^{M}, \alpha_{rj}^{M}, \gamma_{rj}^{M}, \delta_{ij}^{M}$ and $\beta_{ij}^{m}, \alpha_{rj}^{m}, \gamma_{rj}^{m}, \delta_{ij}^{m}$ as the maximum and minimum optimal input/output weights respectively, as obtained from M6.1. These values correspond to the maximum or minimum weight that each input/output can take without violating the efficiency of the DMU concerned and will be used to modify the formulation of the DTP model in M6.4. The mathematical model for estimating the minimum and maximum optimal weights is given in Appendix 6A.

The extra information, concerning the range of optimal values for inputs/outputs, can be incorporated within the formulation of the DTP in the set of equations in M6.4b. **This modification will lead into the non-linear form in M6.5 where the weights and the quantities of inputs and outputs become variables of the model simultaneously.**

---

[1] In a resource allocation case some production factors can have a variable treatment. For instance, the demand for services in the public sector is in principle fixed; however, a reorganisation policy would examine the appropriateness of re-allocating demand from least efficient to more efficient operating units.

$$\begin{array}{lll}
\sum_{i \in I_c} \beta_{ij}\phi_{ij} - \sum_{r \in O_c} \alpha_{rj}\psi_{rj} - \sum_{i \in I_f} \zeta_{ij}x_{ij} + \sum_{r \in O_f} \gamma_{rj}y_{rj} - t_j = 0 & & \\
\beta_{ij}^{m} & \leq \beta_{ij} \leq & \beta_{ij}^{M} \\
\alpha_{r}^{m} & \leq \alpha_{r} \leq & \alpha_{r}^{M} \\
\zeta_{ij}^{m} & \leq \zeta_{ij} \leq & \zeta_{ij}^{M} \\
\gamma_{rj}^{m} & \leq \gamma_{rj} \leq & \gamma_{rj}^{M}, \\
\beta_{ij}, \alpha_{rj} \text{ free variables, } \zeta_{ij}, \gamma_{rj} \geq 0, & & \\
\phi_{ij}, \psi_{rj} \geq 0. & &
\end{array} \tag{M6.5}$$

The formulation in M6.5 yields a non-linear problem (in the constraints) as both the β_{ij}, α_{rj} and the ϕ_{ij}, ψ_{rj} terms are unknown. The non-linear terms $\beta_{ij}\phi_{ij}$ and $\alpha_{rj}\psi_{rj}$ can be linearised by introducing variables, $B_{ij} = \beta_{ij}\phi_{ij}$ and $A_{rj} = \alpha_{rj}\psi_{rj}$, and the equations in M6.5 will be converted as in M6.5a.

$$\begin{array}{lll}
\sum_{i \in I_c} B_{ij} - \sum_{r \in O_c} A_{rj} - \sum_{i \in I_f} \zeta_{ij}x_{ij} + \sum_{r \in O_f} \gamma_{rj}y_{rj} - t_j = 0 & & \\
\beta_{ij}^{m}\phi_{ij} & \leq B_{ij} \leq & \beta_{ij}^{M}\phi_{ij} \\
\alpha_{rj}^{m}\psi_{rj} & \leq A_{rj} \leq & \alpha_{rj}^{M}\psi_{rj} \\
\zeta_{ij}^{m} & \leq \zeta_{ij} \leq & \zeta_{ij}^{M} \\
\gamma_{rj}^{m} & \leq \gamma_{rj} \leq & \gamma_{rj}^{M}, \\
B_{ij}, A_{rj}\ \zeta_{ij}, \gamma_{rj} \geq 0, & & \\
\phi_{ij}, \psi_{rj} \geq 0. & &
\end{array} \tag{M6.5a}$$

The use of M6.5a in the formulation of DTP model will proceed the allocation of resources to DMUs without being affected by the presence of multiple optimal sets of weights for efficient DMUs. The representation of efficient units using **ranges of plausible optimal weights** in the solution of M6.4. leaves open the judgmental question as to which set of optimal weights should be selected to represent efficient DMUs. This question is inextricably linked with the incorporation of value judgements by decision makers in the assessment of targets, but it has not been pursued any further in the thesis. It is noteworthy, however, that the solution to the modified model M6.4 will yield input/output values for the efficient DMUs as they are represented in the remaining objectives, notably effectiveness and equity.

3.3. Defining equity goals in the decentralised planning model

The set of constraints in M6.4c provides a quantitative representation of equity. The constraints are based on the proportional representation of the relative need RN_j of DMU j as was defined in M6.3. The allocation of controllable inputs ϕ_{ij} in the planning model is represented in M6.4c as $\phi_{ij} - \varepsilon_{ij}^{+} + \varepsilon_{ij}^{-} = Gx_i * RN_j, \forall i, j$ where Gx_i is the global level of the i^{th} controllable input, RN_j is the relative need of unit j and $\varepsilon_{ij}^{+}, \varepsilon_{ij}^{-}$ are goal deviation variables representing over/under achievement of the equitable allocation of input i for DMU j.

Mandell (1991) and Savas (1978) argue that the exact representation of equity in allocating resources and/or public goods/services can vary. This variation is a function of different assumptions concerning the nature and measurement of equity. The numerical representation of equity has been considered so far via the assessment of *relative need* in M6.3 which is a computationally tractable process based on multi-attribute value functions. Equity can also be defined using the concept of *inequality* which has, perhaps, more theoretical elegance at the expense of incorporating computational complexities in the resource allocation process (multi-objective nature). The use of inequality measures (e.g. Gini coefficient, Atkinson (1970)) as a means for representing equity, and its implications on the formulation of the DTP model are discussed in more detail in Appendix 6B.

3.4. Incorporate policy constraints in the planning model

The last set of constraints, notably M6.4d, seek to accommodate policy issues within the planning model. The upper and lower bounds of the allocated inputs/outputs can be used to facilitate the use of legislative type of constraints. For example, the allocation of staff to a school would include staff-student ratio, student per classroom and staff with different experience/grades as policy constraints that need to be enforced by the planning system.

Policy constraints can also be used to restrict the radical reallocation of resources among DMUs that could be pursued otherwise. This can increase the operational feasibility of the solution to M6.4 by minimising the *managerial* and *political turbulence* that follow a radical reallocation system implemented in a *not-for-profit* MUO.

Finally, the upper and lower bounds can be used as an instrument for bringing the DTP to the budgeting/planning process followed by organisations. For example, upper and lower bounds can be used to incorporate the budgeting proposals of individual DMUs (upper bounds for resources and lower bounds for output) and central management (lower bounds for resources and upper bounds for outputs) prior to the allocation process.

3.5. The objective function of DTP as a planning instrument

Having discussed the representation of resource planning objectives (effectiveness, efficiency and equity) in DTP, emphasis is next given on the form of the objective function in M6.4. The objective function is made of three sets of deviation variables which are associated with the corresponding objectives of resource allocation. Note here that the objective function in M6.4 is currently presented as a three phase goal programming problem by not linking the three sets of deviation variables in the same utility function. This issue is discussed in more detail in the next section of the chapter.

- **Effectiveness** is represented by the global target levels in M6.4a. The deviation variables $\sum_{r\in O_c}\frac{P_r^+D_r^+ + P_r^-D_r^-}{Gy_r}+\sum_{i\in I_c}\frac{P_i^+D_i^+ + P_i^-D_i^-}{Gx_i}$ included in the objective function have been standardised per unit of input/output correspondingly, to avoid scale bias in the final solution of M6.4. The model also includes preferences related to per unit penalties for not achieving the level of assessed global targets.

- The second group of goal variables concerns the conservation flow equations and the minimisation of the associated **inefficiency** components $\sum_{j=1}^{n} P_j t_j$. The preferences attached to each inefficiency component P_j give the option for expressing different priorities over the elimination of the inefficiency of individual units.

- Finally, the third part of the objective function relates to the relative need and policy constraints which seeks to incorporate equity as a planning objective. The constraints associated with ***equity*** are pure goal constraints and, therefore, under and over achievement deviation variables are needed. These deviation variables $\sum_{i\in I_c}\sum_j\left(P_{ij}^{\varepsilon-}\varepsilon_{ij}^- + P_{ij}^{\varepsilon+}\varepsilon_{ij}^+\right)$ are included in the objective function of the DTP. The trade-off between the two-side deviation variables is expressed in the preference levels associated with each deviation variable. A judicious choice of these preferences could promote redistribution of resources to classes of DMUs with particular characteristics.

The weights of preference attached to the deviation variables in the objective function capture internal tradeoffs **within** each of the three objectives, namely effectiveness

(P_r^+, P_i^-), efficiency (P_j) and equity $(P_{ij}^{\varepsilon-}, P_{ij}^{\varepsilon+})$. The case of external tradeoffs **between** the three objective functions needs to be addressed separately in the solution process. The reasoning for separating the process of expressing priorities **within** the attributes of each objective and **between** the objectives themselves is twofold.

- Firstly, the decision makers will not be distracted by compounding issues related with the representation of an objective, and issues related with the tradeoffs between objectives,
- Secondly, the priorities concerning the internal structure of each objective can be articulated *a-priori* by decision makers, while the tradeoffs between the three objectives are pursued interactively.

These issues are discussed in more detail below in the operationalisation of the solution process of the DTP model.

4. Operationalising the decentralised target-based planning model

The modelling phase of the decentralised planning model was, hitherto, based on the mathematical representation of alternative objectives of planning under a goal programming framework. The remainder of the chapter focuses on issues concerned with the application and solution process of the decentralised planning model. This includes a discussion on the input/output variables, an algorithmic process for the implementation and selection of the solution method and finally, an investigation of the conditions of efficiency and optimality which concerns the solutions obtained by DTP.

4.1. Selection of input-output variables

M6.4 is made up of mathematical statements between those inputs-outputs that constitute the production process of DMUs and thereby the global organisation. Moreover, the model seeks to reflect the structure of the organisation concerned and, therefore, anticipates the presence of decision makers at different levels of management. This has two main effects upon the nature of the input-output variables employed.

The **first** concerns the degree of input/output controllability by different levels of management. Thus far the formulation provided in M6.4 has adopted an assumption that the inputs/outputs used in the DTP are uniformly controllable within the organisation. This assumption was made mainly for simplicity and it can be relaxed. This would advance the formulation in M6.4 by introducing variable input/output controllability between different levels of decision making and, furthermore, between DMUs. This can be pursued by using the concepts of variable degree of input/output controllability introduced in chapter four (see M4.3).

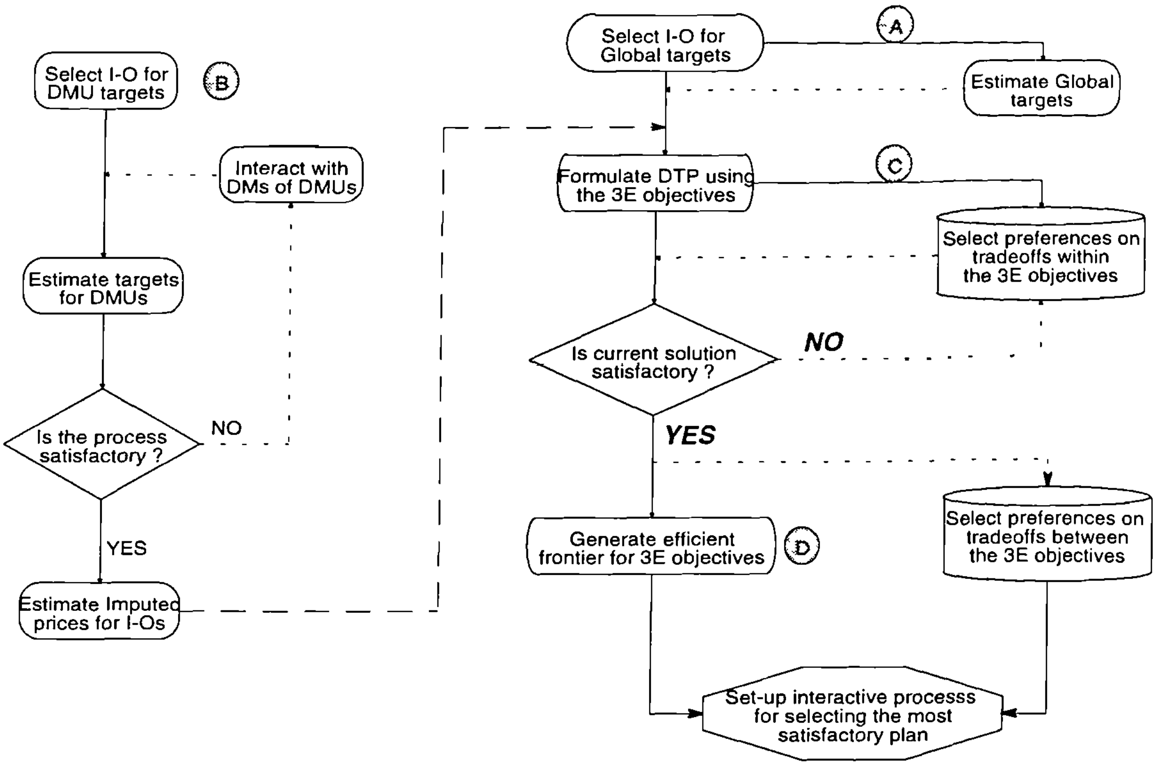
The **second** relates to the choice of inputs/outputs in the model. There is an assumption that the organisation as a whole can decide on commonly accepted inputs/outputs. However, the selection of inputs/outputs that will be acceptable by all levels of decision making is not given *per se*. For the central government, for example, the allocation of central grants to local authorities is a variable to be minimised while, on the other hand, local authorities seek to maximise the level of the same variable. A decentralised planning model needs to address these conflicting views within the formulation and the solution process.

4.2. Solution process of DTP

An algorithmic representation of the decentralised planning model is provided next focusing on the requirements of implementing DTP in a real life organisation. From the mathematical development of the method it was obvious that the decentralised model would require information obtained from independent solution stages. The flowchart in Figure 6.3 represents the various components required for the development of the decentralised target-based planning model.

Figure 6.3

Decision support for decentralised target-based planning



The goal programming model in M6.4 seeks to integrate the three determinants of resource allocation, namely *effectiveness, efficiency, and equity* under *a* decentralised target-based framework. The decentralised system, however, is facilitated via the representation of individual DMUs carrying their own "technology" to the planning model. **The degree of participation of different management levels in the planning process of multi-level organisations is a key determinant of decentralisation, Brooke (1984).** Lack of absolute domination from central management, however, reinforces the need for well defined co-ordination mechanisms that will guarantee the representation of all interested parties, as well as the system's functionality. The flow chart concerning the structure of the decentralised target-based planning model in M6.4 includes four interrelated phases. These phases are discussed next with emphasis on the representation and co-ordination of decision makers' preferences over the satisfaction of the planning objectives.

- **PHASE A AND PHASE B:**
  Trade-offs among different levels of decision making,

The process is initiated in phase A by assessing global performance targets Gx_i and Gy_r in M6.4. The central management is responsible for this process and the global targets are estimated following the methods discussed in this chapter. The second phase of the process focuses on the assessment of performance targets for individual DMUs using M6.1. This process yields the technological coefficients for inputs and outputs that are used to build up the efficiency equations M6.4b in the DTP model. The estimation of these technological coefficients will incorporate preferences from lower level management concerning the relative importance of inputs/outputs in assessing targets.

Phase A and phase B of the solution process initiate the decentralising debate by allowing different management tiers to influence the planning process of the organisation. The different management tiers involved in the two phases of planning may differ on the definition of the input/output production sets, the degree of controllability of the inputs/outputs and finally the relative importance given to the improvement of different the inputs/outputs.

- **PHASE C :**
  Trade-offs within the planning objectives

Phase C seeks to synthesise the results obtained in phases A and B and proceeds in formulating the goal programming model in M6.4. Each of the three planning objectives represented in M6.4 have composite nature encapsulating a set of secondary objectives (attributes). It is vital, therefore, at this stage to proceed towards reflecting upon the **internal tradeoffs** within each planning objective.

Effectiveness for instance relates to the achievement of global targets of controllable inputs/outputs. Central management would, therefore, express a preference structure ($P_r^{+,-}, \forall r$ and $P_i^{+,-}, \forall i$) concerning its interest for achieving global targets for each input or output variable. Similarly, the representation of efficiency as a planning objective would include different preferences (P_j) for minimising the inefficiency of clusters of DMUs. Finally, when equity is represented in the DTP model, decision makers are allowed to express different priorities ($P_{ij}^{\varepsilon-,+}, \forall i \in I_c$) between either the importance of the different controllable inputs to be allocated equitably or the importance of different DMUs of the organisation to be resourced equitably.

The priorities concerning the internal tradeoffs for each the three planning objectives can be derived using methods from the relevant literature. That is to say, techniques which include the Analytical Hierarchy Process (AHP), the Centroid method, the Pair comparison method can be used to facilitate decision makers in order to quantify their preferences concerning the importance of attributes in the interior of each objective in the model. The use of these methodologies for incorporating preferences *a-priori* among the attributes of each objective in goal programming are advocated *inter alia* by Gass (1986), Srinivasan (1973) and Sinuany-Stern (1984).

Having estimated the weights of importance for the attributes of individual objectives one can proceed to the final part of the solution process in phase D.

- **PHASE D:**
  Trade-offs between the planning objectives

As discussed earlier on, model M6.8 has a multiple objective nature. The tradeoffs between the three planning objectives, namely *effectiveness, efficiency and equity,* need to be addressed by synthesising the three objective functions in M6.4a. Each objective function includes priority weights (see phase C) for its attributes without, however, considering the conflicts between the simultaneous maximisation of the three objectives. The development of a two-stage process for accommodating the conflicting character **among** the objectives' attributes and **between** the objectives themselves is not an issue discussed in the current goal programming literature.

In the context of the decentralised target-based planning model the acknowledgement of the two level of tradeoffs is a key issue of the model. In particular, the tradeoffs between the three objectives constitutes an issue with profound economic and policy implications in the decision making process of *not-for-profit* organisations. The potentially conflicting relationship between these three objectives is well recognised in the literature. The formulation of the DTP model, however, advances this debate significantly since the three objectives coexist within the objective function of the same optimisation problem. It can be argued that this objective function encapsulates all elements of the welfare of the organisations concerned.

The tradeoffs between the three planning objectives can be addressed operationally using an interactive goal programming solution approach. Reviews of the interactive goal programming methods can be found in Zionts and Wallenious (1976), Zeleny (1982), Goicoechea *et al* (1982), Hwang and Masoud (1979).

The review of advantages and disadvantages of interactive multi-objective programming methods is beyond the scope of this research as the selection of appropriate interactive method is mainly context dependent. In other words, the size of the problem, the number of decision makers involved, and the organisational structure can all often determine the choice of interactive mechanism to be employed.

4.3. Efficiency and optimality in the solution of DTP

The formulation of M6.4 is a goal programming one. Goal programming is seen as one of the most successful and widely used multi-objective programming methods. The method has been widely applied, Zanakis (1982) and Romero (1991), in many areas of business and economics. However, a number of concerns have been raised regarding the limitations of the method. These discussions have been the subject of major academic debate between Zeleny (1982), Hannan (1981), Romero (1991) and Ignizio (1982). More recently Min and Storbeck (1991) summarised pros and cons of goal programming in an attempt to disentangle accumulated misconceptions of the use of the method.

The main body of the chapter concentrates on the properties of the solutions obtained from M6.4, whilst Appendix 6C provides some mathematical tests that would allow to investigate these properties.

The goal programming model in M6.4 can take the form of a medium to large scale mathematical programming problem. As Kornbluth and Salkin (1987) argue large, scale problems very often suffer from the presence of multiple optimal solutions. This, however, brings forward the question on whether the solution obtained from M6.4 is *Pareto efficient.*

> Pareto efficiency is a fundamental concept of multiple objective programming. Given a set of objective functions $F_p(\mathbf{x})$, p=1,...,n taking values from a convex solution space $\mathbf{x} \in X$ the concept of efficiency is defined as follows. A solution vector $\mathbf{x}$ is efficient if $F_p(\mathbf{x}) \geq F_p(\mathbf{x}')$, $\forall\ \mathbf{x}' \in X$ and $\forall$ p=1,...,n with at least one strict inequality holding.

In the case of M6.4 one needs to investigate the possibility of obtaining non-efficient solutions to the problem. This is possible in the goal programming formulations that include under and over achievement deviation variables or in the goal programming problems solved by the lexicographic method, Zeleny (1982). A number of different methodologies are suggested in the literature, for investigating the presence of inferior solutions in goal programming problems. All of these methodologies include further computations in the original goal programming model, see Hannan (1981), Min and Storbeck (1991) and Romero (1991).

A goal programming model which investigates the presence of inferior solutions in M6.6 is provided in Appendix 6C. Since the computation process of goal programming problems is not the main concern of this chapter the models for generating non-inferior solutions to goal programming models are provided for completeness. Undoubtedly, a numerical application of decentralised planning in future research would need to further investigate the properties of the solutions obtained from M6.4.

4.4. Decentralised planning model & the principles of target setting

The DTP model shares the same aims with the CTP model developed in chapter 5 insofar as the development of performance based resource allocation systems are concerned. A key part of the development of the DTP model concerned its ability of direct accommodation of the objectives of resource management namely effectiveness, efficiency and equity. This direct representation is not feasible under the formulation of the CTP model in chapter 5. The goal programming formulation in M6.4 makes explicit reference to each of these objectives including their quantitative representation. It is noteworthy that the DTP model in based on the classical use of linear programming models for allocating resources. The crucial element of the formulation in M6.4 is based on the estimation of the "technological coefficients" which link the inputs/outputs of individual DMUs.

With regard to the principles of target setting introduced in chapter 3 the following can be argued.

The competing nature of efficiency and equity as objectives of resource management is well recognised amongst economic and political science literature. In both cases, however, the arguments are theoretically led without real concern to develop operational models. Attempts to address these issues at the operational level (see Mandell (1991)) were based on *a-priori* assumptions regarding the theoretical <u>efficiency</u> and <u>effectiveness</u> association between inputs and outputs. The formulation of the DTP model has avoided this type of assumption as it uses empirical estimates of relative efficiency and effectiveness. This is in direct accordance with the principle of *value revelation* advocated earlier in chapter three.

Another important feature of the DTP model is the multi-level multi-stage process for its implementation. Recall that the estimation process of the technological coefficients for the inputs/outputs via M6.1 seeks to encompass the preferences of low level management. On the other hand, the preferences used in the solution process of M6.4 are driven by the aims of the central management which is responsible for the allocation of resources. Clearly, the DTP model has *multi-level* dimension which is further advanced by the representation of the *global organisational targets* as quantitative goals to be achieved.

The decision variables (ϕ_{ij}, ψ_{rj}) of M6.4 correspond to controllable inputs and outputs of individual DMUs which make explicit reference to the resource allocation nature of the problem. The allowance of *resource transferability* among individual DMUs reinforces the presence of *equity* as an objective of resource allocation. The *decision support* process implemented via the solution of the problem has an interactive nature which aids the achievement of *managerialy feasible solutions.*

Finally, the explicit reference made during the solution of M6.4 to the tradeoffs **between** different levels of management, **among** the criteria of each objective and **between** the achievement of the three objectives, enhances the policy making and decision support role of the decentralised model in not-for-profit MULOs. These issues are pursued by incorporating the decision makers' preferences throughout the whole development process of the DTP. This is a element of strength of the DTP model as by "letting the man in" (see Zeleny (1992)) increases the likelihood for obtaining scenarios to be implemented.

5. Conclusions

In this chapter the decentralised target-based planning model was put forward as a means for linking managerial control and planning in MUOs. The central point of this linkage was once more the joint assessment of DMU and global based organisational targets. A similar problem was also addressed in the previous chapter (chapter 5). The uniqueness of the methods developed in this chapter emanates from the adoption of decentralised decision support for target setting and resource planning. Particular emphasis was given on the representation of individual DMUs by the technological coefficients that reflect the aspirations of management from the same rank within the organisation. Undoubtedly this approach is computationally expensive as the formulation of the planning process uses components that are obtained from solutions to unrelated performance measurement problems.

The explicit representation of the three planning objectives in MUOs within the planning model gives merit to the DTP model for being a reliable planning tool. Interactive/iterative processes need to be developed, however, in order to aid DMs in selecting the most satisfactory planning scenario.

The decentralised planning model DTP is a more demanding decision support tool by comparison with its centralised counterpart of chapter 5. It is anticipated, however, that the plans obtained via decentralised decision making processes have a higher likelihood of being implemented as their development has received a higher degree of consensus among the hierarchical levels of multi-unit multi-level organisations.

- END OF CHAPTER SIX -

Chapter 7

Diagnostic analysis and decision support for a-priori resource allocation[1]

1. Introduction-motivation

Chapter 7 begins the investigation of the development of *a-priori* decision support systems for managerial control and planning. The distinction between *a-posteriori* (chapters three to six) and *a-priori* resource allocation has been discussed in the first chapter of the thesis. Both methods have general applicability but the *a-posteriori* method is more suitable for not-for-profit MUOs whilst the *a-priori* method is geared more towards for-profit MUOs. This chapter pursues diagnostic issues regarding the assessment of the performance of individual DMUs whilst chapters eight and nine are linked to the planning issues of the *a-priori* decision making.

A basic characteristic of *for-profit* organisations is undoubtedly the competitive (market) environment in which they operate. Market oriented MUOs operate under intense competition and therefore "operating efficiency" is a key success factor that affects their long run viability. The definition of operating efficiency, however, is not directly linked with the actual profit generated by individual firms. Based on Farrell's (1957) ideas of technical efficiency this chapter defines the concept of operating efficiency in profit making MUOs and proposes models for its assessment.

[1] A version of this chapter has appeared in the Journal of the Operational Research Society (1995), Vol. 46, Issue 1.

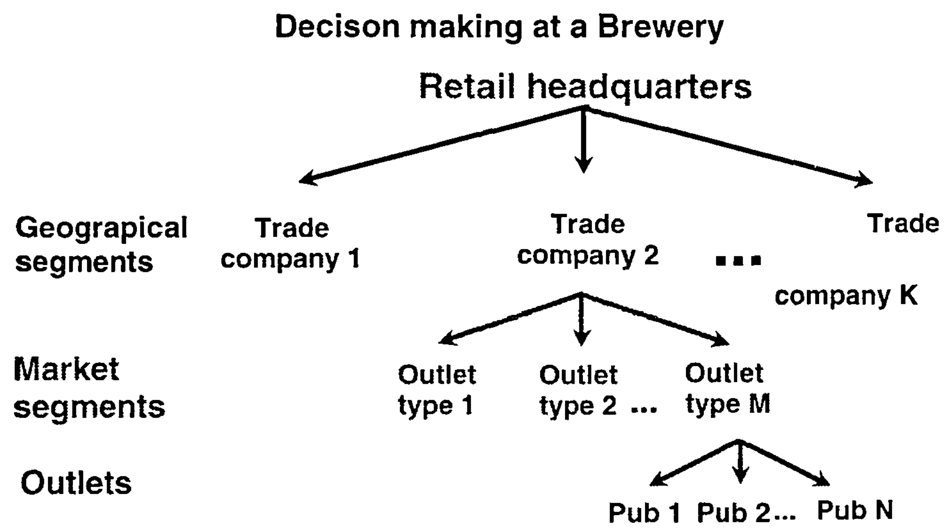
Decision making in *for-profit* MUOs is concerned with the best deployment of resources within networks of DMUs in order to achieve the corporate objectives of the organisation. The achievement of corporate objectives would give competitive advantage to individual firms which in turn affect their long run viability. The management of MUOs is often called upon to make decisions concerning

- the development of innovative products/services in order to create/utilise market opportunities,
- to identify effective location strategies for opening new outlets,
- to explore the market changes in the surrounding area of existing outlets,
- to develop benchmarking strategies for effective managerial practices among DMUs,
- to identify effective product mix services for different markets and promote marketing policies for supporting their products and services,
- to allocate capital for maintaining and improving the profile of the network of DMUs (e.g. training, redecoration, expansion).

The issues listed above largely describe practices followed by a wide range of profit making MUOs. These issues can be addressed effectively only if an organisation combines the decision making responsibilities of different levels of management. The latter brings forward the multi-level character of decision making in order to support *a-priori* resource allocation decisions. The operation of a large retailing brewery can be used to illustrate the nature of multi-level multi-unit operations of for-profit MUOs. Data from public houses of this brewery will also be used to obtain empirical results of models developed in this and the subsequent two chapters. Figure 7.1 describes the levels of decision making that are used in the operations of a retailing brewery.

Figure 7.1

Decision & control in multi level profit making organisations



The hierarchy displayed in Figure 7.1 contains three levels of decision making. Each level of decision making has a different degree of authority regarding the allocation of resources and the assessment of performance. The multi-tier decision making and assessment of performance implies that decision support models should be customised to reflect different levels of management.

In the brewery, for example, individual pubs make proposals for investments, the trade companies prioritise and select part of these proposals within their investment portfolio and finally the central organisation is responsible for allocating capital among the trade companies for realising their investment plants. The distinction between the levels of decision making concerns: the central management who is responsible for the performance of the organisation as a whole; middle management who are responsible for planning decisions and, local management who are responsible for the operation of individual units.

Provision of decision support has *diagnostic* and *planning* stages. ***Diagnostic*** **analysis concerns the assessment of units' performance based on decisions and resource commitments made in the past. The analysis, therefore, focuses on the extent to which individual DMUs responded in support to the global organisational performance.** ***Planning*** **analysis has broader appeal and seeks to investigate the appropriateness of decisions that would result in future resource commitments (e.g.**

capital investment, reorganisation policies, etc.). Undoubtedly, diagnostic and planning analysis are inextricably linked decision support tools for MUOs.

For-profit DMUs are typically assigned geographical segments and seek to penetrate the markets within these segments by generating sales. Of equal importance is the conversion of sales into profits. The two objectives will be discussed in more detail later in the chapter but for the moment the discussion concentrates on defining the domain of *diagnostic* and *planning* investigations in MUOs.

Figure 7.2 below is used to facilitate this discussion.

Figure 7.2
Diagnosis & Planning issues in profit MUOs

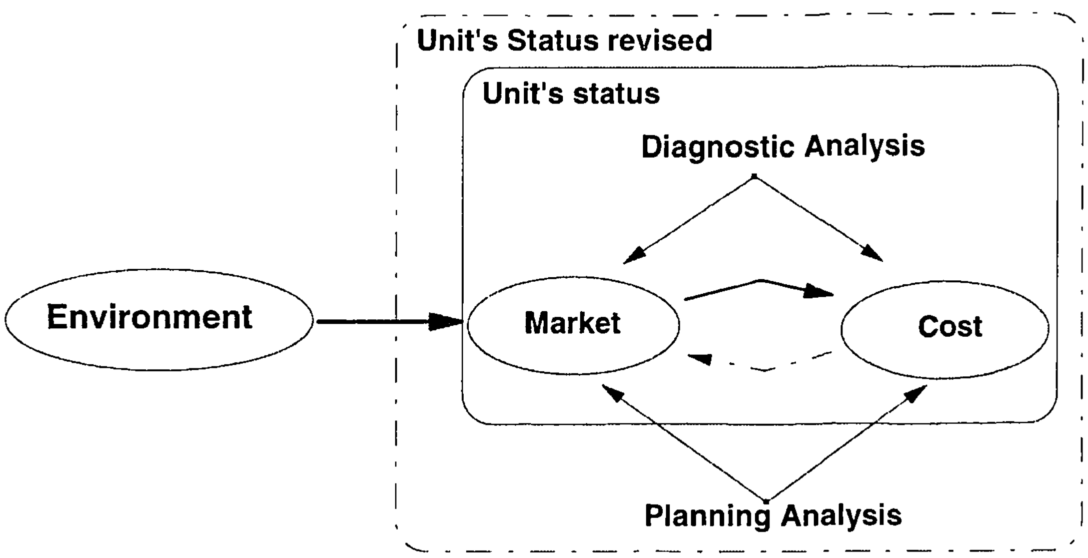


Figure 7.2 shows diagramaticaly the role of diagnostic and planning analysis in MUOs as part of the process for controlling the performance dimensions of individual outlets. Diagnostic analysis concerns the extent to which individual units perform satisfactorily given their objectives and resources (inner domain). Planning analysis seeks to explore the scenarios for improving units' performance by changing their objectives and resource base (revised domain). **The development of systematic diagnostic analysis procedures in MUOs is the main theme of this chapter.** The implementation of diagnostic analysis will be made focusing on:

- the definition of the determinants of performance in profit making MUOs,
- the assessment of performance targets for different tiers of management based on the previously defined performance components,
- the identification of good/exemplary operating practices that will be used as the benchmarks in MUOs.

The rest of the chapter is organised as follows:

Market and *cost* efficiency are defined as the main components of the performance of profit making MUOs. Frontier analysis is used for assessing cost and market efficiency of profit making units. The abstract concepts of market and cost efficiency will be conceptualised further with a case study on the public houses of a brewery in the UK. Frontier analysis models will be developed and empirical results will be obtained for assessing outlets' performance.

2. Concepts of performance in profit-making MUOs

Profit generation is a goal *for-profit* making units. This needs to be reflected in performance measurement mechanisms employed by such MUOs.

Measures of profitability can be absolute, relative or both. Absolute measures of profitability are "Profit before Interest and Tax" and "Net Profit after Interest and Tax". Relative measures of profitability relate the absolute profit to the revenue from which it was generated or to the capital employed by the unit. Typical relative measures of profitability are the "Gross Profit Margin", representing the percentage of gross revenue that is profit before interest and tax, or "Return on Capital Employed", expressing profit before interest and tax as a percentage of the capital employed by the unit.

Profitability measures, however, are incomplete in at least one important respect. They fail to take into account environmental factors affecting profitability. Such factors need to be taken into account at the very least in the interests of equitable comparisons of units. More importantly, however, they make it possible to assess the viability of units and to identify managerial practices conducive to higher profitability. For example in the case of a shop Norman and Stoker (1991) argue its ability to control its costs (cost efficiency) and its ability to attract custom (market efficiency) are independent aspects of performance. They should therefore be assessed separately if the shop is to be set sensible performance targets.

To illustrate how profit figures alone may be inadequate to convey the performance of a unit consider the assessment of a set of bank branches. Environmental factors such as the income levels, age distributions and level of competition in the catchment area of each branch are outside the control of a branch yet they influence its profits. A very profitable branch may in fact be foregoing even higher profits given the environment in which it operates. Conversely, a branch with low profits may in fact be doing very well for the environment in which it operates. Its practices could be very effective in generating high profits if employed in a more helpful environment.

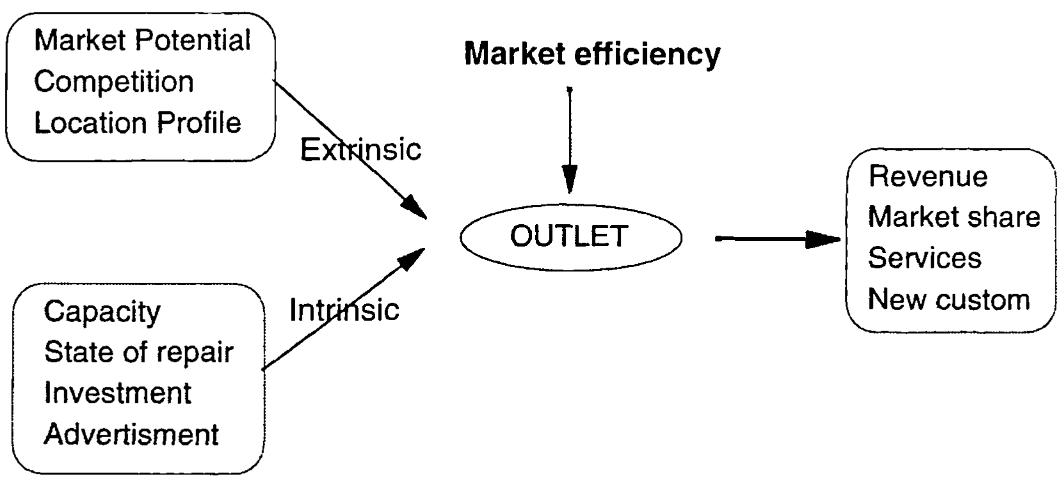
Market, cost efficiency and profitability constitute key performance components of profit-making DMUs. The nature of these components is illustrated next using Figure 7.3 and Figure 7.4.

2.1. The notion of market efficiency

Market efficiency of a DMU is the extent to which it penetrates its own market as compared with other DMU operating similar functions. This is assessed on the basis of output attributes expressed in either monetary terms i.e. revenue, market share, or as pure service/volume quantities, i.e. number of transactions with clients, number of new clients, volume of quantities produced or sold, etc. DMUs' market efficiency is affected by groups of extrinsic and intrinsic input attributes. Some of these attributes are listed in Figure 7.3.

Figure 7.3

Extrinsic & intrinsic attributes of market efficiency



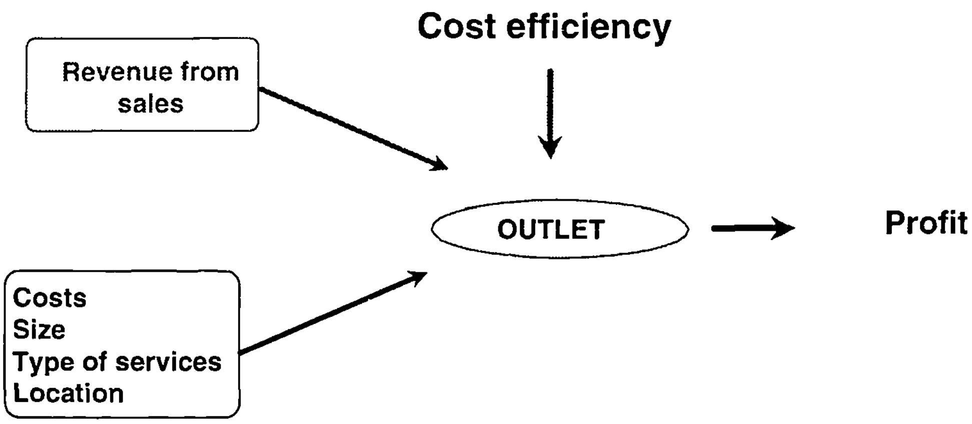
Intrinsic input attributes reflect outlets' marketing profile determined by factors such as capacity, state of repair, age, staff quality, capital investment, advertising, etc. These are the only factors over which management[2] has some control. **Extrinsic** factors on the other hand reflect market characteristics of the outlets' surrounding environment. These factors are primarily uncontrollable as they relate to consumer behaviour, population make-up, competition, location characteristics (outlet's visibility), etc. The joint use of intrinsic and extrinsic input attributes allows management to assess the extent to which outlets achieve the highest possible results in terms of revenue.

With reference to measuring market efficiency units can be seen as operating a "production technology" in which inputs are the environment and the resources deployed by the unit while output is the revenue generated. The efficiency of a unit within this production technology is its market efficiency. **Thus, market efficiency reflects the units' ability to convert potential for sales into actual sales.**

2.2. The notion of cost efficiency and profitability

Cost efficiency is an internal factor in the operation of outlets. Depending on the nature of the operations of DMUs cost efficiency can be a very important determinant of performance. This would apply to cases where there is little scope for improving market efficiency of outlets and therefore any performance improvements coincide with improved cost control. This type of situation often arises in profit making DMUs like patrol stations where the demand for services is predetermined by the location of the store and the prices of petrol. Profit generation can be thought of as an internal process of a unit as illustrated in Figure 7.4.

Figure 7.4
Attributes of cost efficiency



---

2 It should be noted however, that intrinsic input attributes are under the discretion of central and/or local management.

Cost efficiency is defined as the ability of outlets to convert their revenue into actual profit. To assess cost efficiency one needs to consider all cost components that "consume" parts of the generated turnover. Factors that affect the operating costs of outlets need also to be considered. These include the size, state of repair, location and the service mix provided by individual units. The state of repair and maintenance costs of a public house with catering facilities, for example, is expected to be higher than for public houses which offer only drink services.

Measures of relative profitability of outlets, defined as profit margin, are often used as surrogates of cost efficiency without however coinciding. There are, for example, cost efficient outlets with very low profitability due to their high fixed costs and/or low market efficiency. These phenomena need to be taken into account when decisions are made concerning the long run viability of individual outlets.

2.3. Frontier analysis for assessing performance of profit making MUOs

Performance issues in profit making MUOs are discussed in the marketing, accounting, and management science literature. In each discipline, however, the motives for assessing performance are different and therefore different performance measures are used. For example, sales forecasting methods for individual outlets are customarily used to support marketing activities and profitability ratios are used by accounting departments to assess the financial performance of outlets and/or outlets' managers.

In summary, sales forecasting and accounting based methods provide exclusive support on either the planning or the diagnostic dimensions of performance. Sales forecasting would guide the opening of new DMUs and would, therefore, be used for assessing "site" performance, whilst accounting profitability ratios would be used to evaluate the profit generation of existing sites. In assessing performance of two bank branches that face different competition, for example, the use of profitability ratios would treat them as if they had the same opportunity for generating sales and subsequently profits. Frontier analysis tools, on the other hand, can encapsulate issues related to the environment in which individual outlets operate and therefore support in full the earlier definitions of market and cost efficiency.

Frontier analysis methods have been used for assessing cost efficiency in financial institutions, e.g. Sherman and Gold (1985), Giokas and Vassiloglou (1990), Giokas (1991), Parkan (1987). Banker and Morey (1986) assess the cost efficiency of restaurants compounding, however, cost factors (e.g. labour costs) with environmental factors (e.g. location characteristics) that affect primarily the market and not the cost efficiency of DMUs.

Less theoretical/empirical studies of frontier analysis can be found in the assessment of market efficiency of profit making DMUs. Athanassopoulos and Thanassoulis (1995) employed frontier analysis methods tailored for assessing market efficiency of public houses. A similar attempt was also made by Mahajan (1991) in assessing the "selling function" of sales forces of an insurance company. Frontier analysis methodologies were used in a decision support mode by Banker and Morey (1993) in order to evaluate the appropriateness of opening new branches of food outlets. Similarly, Athanassopoulos (1993) developed frontier analysis models for estimating targets for reorganising the shape and size of inefficient food outlets.

Research efforts concerning the use of frontier analysis in profit making DMUs have also been made by Charnes and Cooper and their associates (F. Phillips and J. Rousseau), Charnes *et al.* (1991), Charnes *et al.* (1993), Golany *et al.* (1993).

The assessment of the market efficiency in profit making organisations needs to take into account the presence of different levels of management as described earlier in Figure 7.1. This would lead into customising market efficiency in accord with the level of management concerned. Figure 7.5 introduces three types of market efficiency that will be used in this thesis to link organisational structure with the assessment of performance.

Figure 7.5
Multi-level performance measures

Central management	Local management	Planning support
↓	↓	↓
▸ Aggregate market efficiency	▸ Site specific Market Efficiency	▸ Improved Scale Size ▸ Service mix effectiveness
↓	↓	↓
How effective were decisions made in the past ?	*How effective is local management in pursuing its tasks given central decision making ?*	*What can be done in the future regarding the profile of individual outlets ?*

The three components of market efficiency described in Figure 7.5 seek to encapsulate the differences in authority and responsibility among different levels of management in a profit making multi-level organisation. Central management seeks to co-ordinate the overall operations of an organisation by making decisions that affect directly or indirectly all other levels of the managerial hierarchy. Assessment of central management's performance focus on the long run consequences of its decisions which lead to an *aggregate* measure of market efficiency. Local management, on the other hand, is responsible for implementing decisions made by higher level of management. Assessment of local management's performance has a short run horizon as it reflects the *site-specific* market efficiency of individual outlets.

In addition to the two extreme levels of managerial performance, namely *aggregate* and *site-specific,* the multi-level performance measures in Figure 7.5 make reference to a third component which is concerned with issues of decision support. Assessment of performance at these intermediate levels has a decision support role by setting performance targets for individual outlets that would adjust the size and scope of operations of individual outlets. These levels of decision support seek to develop decision scenarios for improving the profile (e.g. scale of operation and service mix) of individual outlets using as a means capital investment and reorganisation policies.

The development of operational models for assessing the components of market efficiency alluded to above is pursued in this chapter and completed in chapter eight. As this chapter concentrates on the development of diagnostic analysis tools within profit making organisations emphasis will be given on the disentaglement between the *aggregate* and *site-specific* market efficiency. The latter will seek to estimate performance targets for inefficient units compatible with central and local management responsibilities. The diagnostic analysis will also concentrate on the identification of exemplary performers among market efficient operating units.

3. The case study

In this part of the chapter the case study is introduced which was chosen to be used as a vehicle for the theoretical developments of diagnosis and planning mechanisms in profit making MUOs. It was felt that the integration of the theoretical models with the case study will give a better understanding of the strengths and weaknesses of the method as a decision support tool.

The case study covers a large retailing firm in the UK which is mainly concerned with brewing and food services world-wide. The analysis includes a set of 154 public houses (pubs) of the brewery located in the northern part of England. These pubs provide food and drink sales and they target local and passing clientele. Background information concerning the brewing industry in the UK can be found in Appendix 7A.

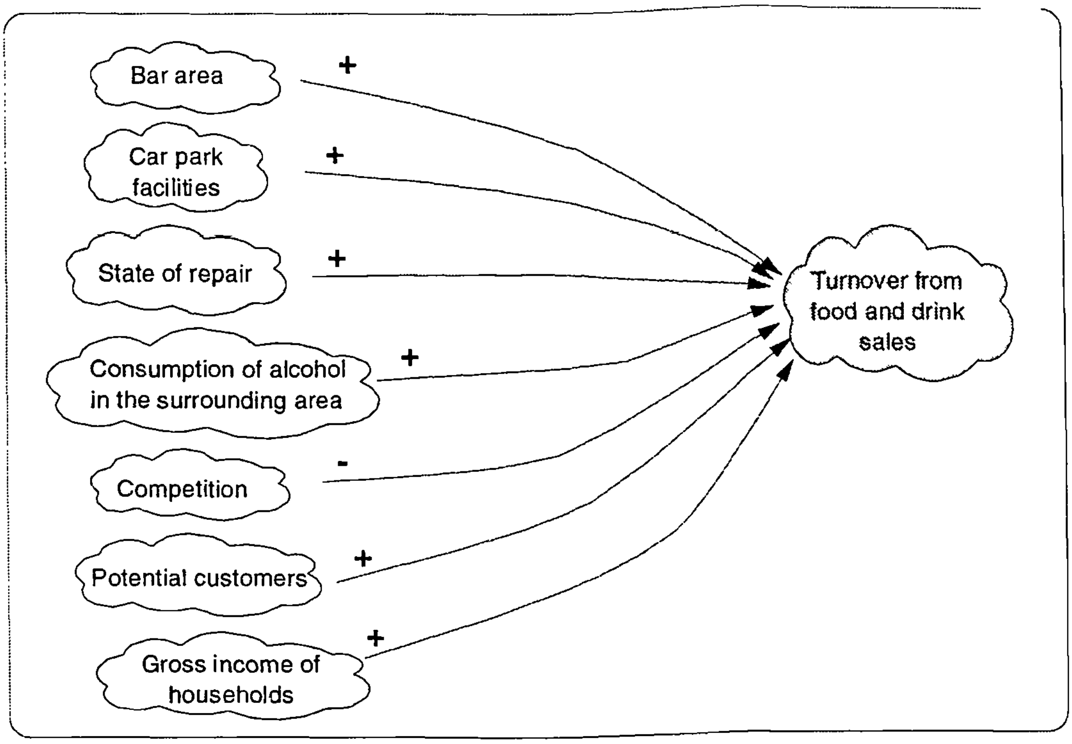
3.1. Ascertaining factors affecting the market efficiency of pubs

Thus far, the concept of market efficiency has been discussed in abstract terms as an important component of the performance of profit making MUOs. This concept will be applied to define the market efficiency of pubs.

In Figure 7.3 it is suggested that the assessment of market efficiency should consider factors representing environmental (market) and internal characteristics of DMUs. The specification of attributes reflecting these characteristics depends on the nature of the MUO assessed. The methodology followed in this study for identifying determinants of market efficiency of pubs was a series of interactions with the central management and the analysts of the Operational Research department of the brewery. Statistical analysis was used to verify "causal" effects that were thought important by the brewery's management. (Results of the statistical analysis can be found in Appendix 7B).

At the end of this process the input-output variables adopted for assessing the market efficiency of individual pubs are exhibited in Figure 7.6.

Figure 7.6
Factors affecting the market efficiency of pubs



The signs attached to each input factor of Figure 7.6 represent the assumption of the causal effect of input factors on the pubs' outputs. For example competition is thought to have a negative impact on the sales of pubs whilst the consumption of alcohol a positive impact on the generation of sales.

The trade area is the geographic area from which the pub draws most of its customers and within which market penetration is highest. Consultation with the planning managers of the brewery revealed that a radius of 2.5 miles was the most appropriate size of trade area for the type of outlets under investigation. More normative approaches for defining trade areas can be found in the location analysis literature, Ghosh and McLafferty (1987). Techniques like the *gravity* models and the *analog* method can be used to obtain more rigorously the trade area of individual outlets.

A discussion of the reason for using each of the factors listed in Figure 7.6 is provided next.

Turnover, **the output in the context of a DEA model**, is the sum of revenues from the sale of drinks and meals. One could support the use of actual volume of sales to allow for price variation of the provided services. This argument, however, does not hold for the pubs in the current analysis as they have uniform prices determined by the central management of the brewery.

The input variables can be classified as *internal* and *external*. Internal variables reflect factors decided upon by the brewery. External variables reflect environmental factors outside the brewery's control. None of the input variables is controllable by pubs' local management.

3.1.1. Internal inputs

These are:

- Bar area (ft^2)
- No. of car park facilities
- State of repair

The *bar area* of a pub reflects its capacity to accommodate customers. Clearly the larger this area the more customers can be accommodated and the larger should be the revenue generated. This can disadvantage pubs which are given by the brewery a large area requiring them to attract a much larger share of their local market to fill up than is the case for most other pubs. However, in the absence of further information it is implicitly assumed that bar area at each pub is not such that it would require a significant proportion of the local market to fill up.

The number of *car park spaces* is an important variable as "broad based" pubs seek passing as well as passing custom and thereby parking facilities may facilitate certain categories of customers.

The *state of repair* of a pub relates to its general decor, furnishings and fittings. This is thought to influence significantly a pub's ability to attract custom.

There is little difficulty in measuring bar area and car park spaces. However, reflecting the state of repair of a pub is difficult. In the case of this study the state of repair of each pub was assessed on a scale from 1 to 15 by the marketing department.

3.1.2. External inputs

The external variables chosen were as follows:

- Number of competitors,
- Number of potential customers,
- Consumption of alcohol in the surrounding area (barrels),
- Gross household income in the surrounding area[3] (£).

Competition affects the ability of pubs to capture a share of the market in their catchment area. It is clearly not a simple matter to reflect competition for a unit. For example in the case of pubs competition may not so much be reflected by the actual count of competing establishments but rather by some measure of their size or of the strength of competition they represent. We had data only of the number of pubs and clubs in the area surrounding the pubs being assessed and this is the variable used to reflect the strength of competition faced by each pub.

A separate issue is whether competition is in fact an input or an output. Competition can be seen as a factor which "consumes resource" within the "production technology" employed by the pub. That is to say effort is required to overcome competition and expand or maintain a certain market share. Competition can also be seen as an input factor in the way capital might be an input in more traditional production technologies. The difference is that the greater the capital employed the higher the output expected while the converse is the case with competition. In both cases, however, they influence the productivity of other inputs. Competition influences the "productivity" of thc market size and the bar area much as capital influences the productivity of labour.

In this assessment it was decided to take competition as an input. This was partly because we found the second argument above more convincing and partly to maintain a single output variable in the assessment.

The input-output model described in Figure 7.6 assumes isotonicity. That it to say for efficient units higher input levels lead to higher output levels. However, the opposite is the case with the number of competing establishments. Therefore, the inverse of the number of

[3] This type of information is usually used on an average per household basis. In this analysis, however, the actual value has been used in order to avoid bias of scale in the assessment of market efficiency that follows.

competing establishments was used as the input variable for competition. Discussions are offered by Charnes *et al.* (1985a) and Golany and Roll (1989) favouring using the 'inverse' approach. In a number of previous applications of DEA (see Norman and Stoker (1991)) in profit making environments no clear indication is made on how competition was included in the analysis. More importantly Mahajan (1991) uses erroneously the number of competing establishments as an input variable which violates the isotonicity assumption of frontier analysis models (for efficient units higher inputs should lead to higher output levels).

Inverting the number of competitors could have been avoided by using instead their difference from some large (subjectively decided) number. The two approaches give generally dissimilar (but not very different) efficiency estimates. For example for our 154 pubs the median market efficiency was 90% when the number of competitors was inverted but only 88% when it was subtracted from 162 (160 being the maximum number of competitors any pub faced in that year).

The higher the consumption of alcohol in the catchment area of a pub, all else being equal, the larger should be its turnover if it is effective in attracting custom. Thus to judge the effectiveness of a pub in attracting custom we must allow for the consumption of alcohol in its catchment area and this explains the use of this variable as an input.

The *potential number of customers* of a pub and *the gross income of households are* used to reflect the size of its market for the sale of meals. Market surveys had found that the typical customer of the pubs assessed was a person aged 15-24 or 35-54 belonging to one of the socio-economic groups B, C1 and C2. (For a definition of socio-economic groups in the UK see Monk (1986)). Based on this information and data from the Office of Population Censuses and Surveys an estimate was made of the number of potential customers in the catchment area of each pub.

It is argued sometimes (see for example Lovell (1993) p.53) that variables such as our external ones, over which pubs have no control, should not be used as input or output variables in the way suggested above. Instead, efficiencies should be estimated in the first instance using controllable input-output variables only. In a second stage, the uncontrollable variables (our external variables) can be used in the context of a regression analysis to explain the efficiency scores obtained. The key difference between this two-stage approach and the one stage approach adopted in the thesis for assessing market efficiency is that in the former inefficiencies are attributed to the external (uncontrollable) variables while in our approach efficiency is measured while controlling for external variables. On balance the one stage approach was selected because it makes it possible to assess how effectively pubs on an individual basis exploit their uncontrollable factors to generate revenue.

4. Managerial tools for diagnostic analysis

The assessment of market efficiency of units and the identification of "best-practice" policies constitute the purpose of diagnostic analysis of profit making MUO. Market efficiency targets will be assessed for inefficient units using the frontier analysis models introduced earlier on. The mathematical basis of these models has already been discussed in chapter 2. The diagnostic analysis will also focus on the performance of those units found relatively efficient in an attempt to disseminate benchmark operating practices.

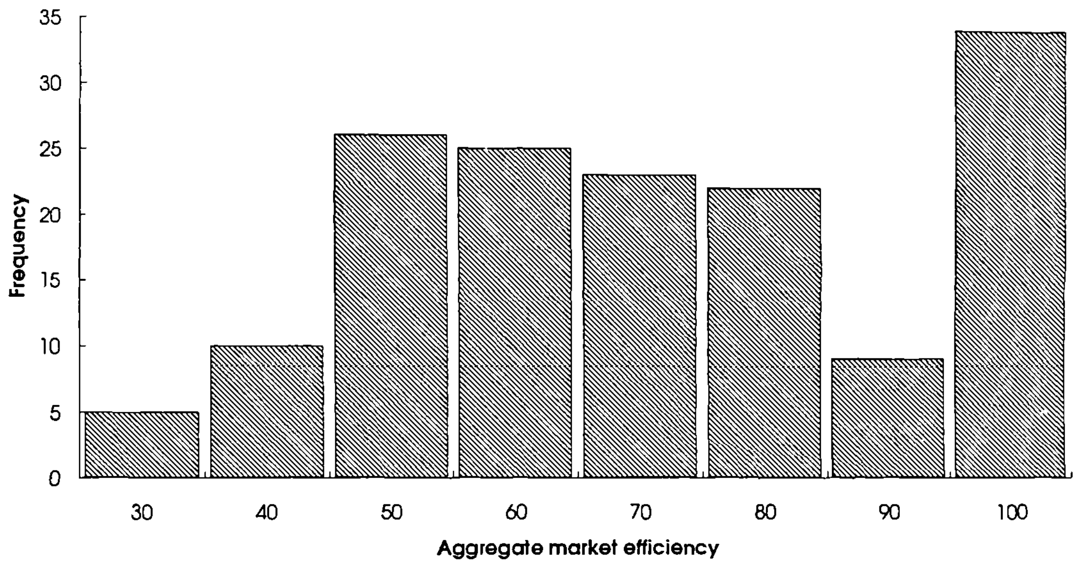
4.1. Assessing aggregate market efficiency

The aggregate market efficiency introduced earlier in Figure 7.5 can be assessed using the frontier analysis models reviewed in chapter two. The input-output set of Figure 7.6 lists a set of i=1,..,7 inputs and the food and drink sales as output that will be used to assess the efficiency of the 154 pubs. In mathematical terms, the market efficiency ($1/z^*$, where $z^*= \max z$), of pub j_O will be assessed using the model in (M7.1)

$$\max z \left| \sum_{j=1}^{154} \lambda_j x_{ij} \leq x_{ij_o} \text{ and } \sum_{j=1}^{154} \lambda_j y_j = z y_{j_o}, \lambda_j \geq 0, i = 1,\dots,7 \ z\,free \right. \qquad \text{(M7.1)}$$

This model is based on the output expansion model described in chapter 2 (model M2.8). The expansion factor z^* gives the proportionate increase to output that would render j_O relatively efficient. This is done by adopting an economic assumption of constant returns to scale in M7.1. That is, DMUs are assessed for their managerial ability to attract custom and also for their scale of operation (input/output mix). **As the scale size of DMUs is mainly under the control of central management the efficiency assessed by M7.1 reflects the aggregate market efficiency of DMUs.** The concept of aggregate market efficiency is illustrated next on the sample of pubs. Results on market efficiency of pubs are summarised in the frequency histogram of Figure 7.7 (market efficiencies stated as a $1/z^*$ percentage).

Figure 7.7
Distribution of aggregate market efficiency of pubs



On average the aggregate market inefficiency is 74% whilst the actual distribution of efficiencies is shown in Figure 7.5. Nearly 30% of the pubs assessed have an aggregate market efficiency below 60% which indicates that high sales' improvements from these pubs should be possible.

The *aggregate market* efficiency compounds factors related with management and the scale of operation of individual pubs. These two factors, namely management and scale, are not controlled by the same level of management. Therefore, the aggregate market efficiency needs to be decomposed by level of management to which it can be attributable.

To investigate further the association between market efficiency and scale size of pubs a chi-square test was employed (see Appendix 7B) which found a significant negative association between aggregate market efficiency and the size of pubs (ft^2). The latter gives some preliminary evidence showing that smaller pubs tend to have higher market efficiency as their turnover (£) per ft^2 of bar area is higher than the corresponding score for the larger pubs (all other input factors being equal). The size characteristics of individual units are under the control of central management and therefore, size free market efficiencies need to be assessed in order to assess the performance of local management.

4.2. Assessing site-specific market efficiency

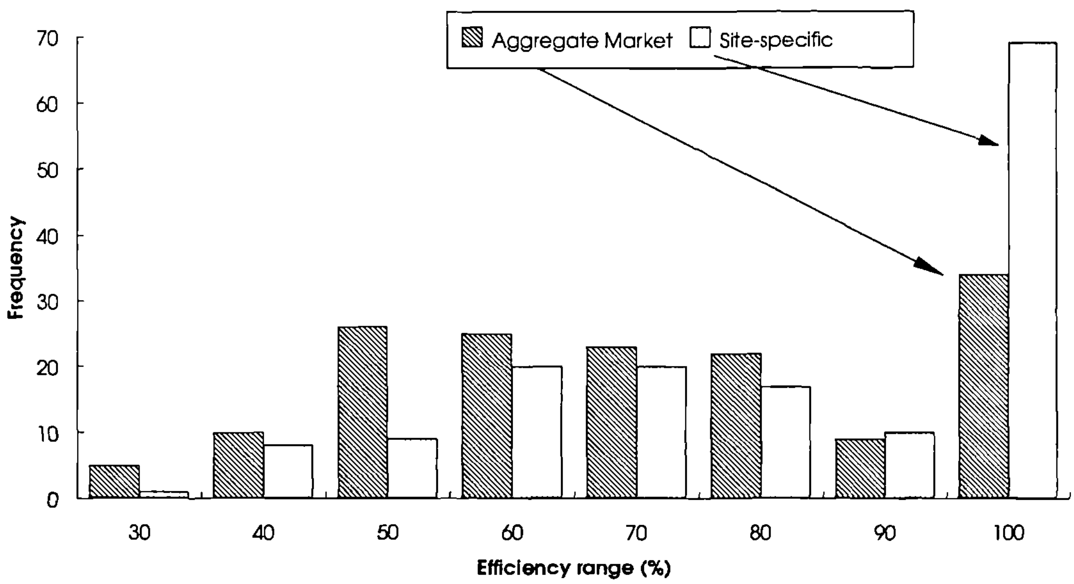
The aggregate market efficiency assessed earlier on concentrates on the central management of an organisation. Central management responsibilities, however, encompass issues related with local management effectiveness and scale size characteristics of individual outlets. The site-specific market efficiency model is employed next to disentangle the aggregate market efficiency into *management* and *scale* components. This can be done using the mathematical model in M7.2 below where market efficiency of units is assessed treating their market size as given (the notation is similar to model M7.1). The expansion factor ρ, that is used to define the **site-specific market efficiency** gives the proportionate improvement in the sales of unit j_O given its scale of operation.

$$\max \rho \left| \sum_{j=1}^{154} \mu_j x_{ij} \le x_{ij_o} \;\vee\; \sum_{j=1}^{154} \mu_j y_j = \rho y_{j_o} \right. \qquad \text{(M7.2)}$$
$$\sum_{j=1}^{n} \mu_j = 1,\ \mu_j \ge 0, i = 1,\ldots,7\ \rho\, free$$

Results on the *site-specific market efficiency* of pubs were obtained using model M7.2 and summarised in Figure 7.8 (market efficiency is expressed in percentage terms as in Figure 7.7).

Figure 7.8

Distribution of aggregate & site-specific market efficiency of pubs



The distributions of aggregate and site-specific market efficiency of pubs are contrasted in Figure 7.8. The average site-specific market efficiency is 83% compared with the 74%

average aggregate market efficiency whilst 12% of the assessed pubs have site-specific market efficiency below 60%. Clearly, the assessment of market efficiency of pubs proves to be highly sensitive on the economic assumptions of returns to scale. The latter shows also the scope of disentangling the potential of performance improvements of pubs among different levels of management.

The assessment of aggregate and site-specific market efficiency of outlets are important components of the diagnostic analysis framework proposed earlier in the chapter. The aggregate market efficiency of outlets was made under the economic assumption of constant returns to scale. The targets estimated by this model have a long run horizon of achievement as they compound management and scale effects on the performance of individual outlets. The bias of scale on performance was relaxed by assessing the site-specific market efficiency which adopts a variable returns to scale economic assumption. Targets obtained by the latter model have more immediate horizon as they refer to the appropriateness of managerial practices employed by individual outlets.

4.3. Identifying role model operating practices (benchmarking)

The diagnostic analysis at the individual DMU level will now focus on those units found relatively efficient in the previous analysis of *aggregate* and *site-specific* market efficiency. The use of DEA does not give performance targets for market efficient DMUs and hence some more qualitative criteria need to be employed in order to give better insights concerning the performance of these DMUs. The process of identifying role model operating practices will be called benchmarking and is discussed in more detail next.

Benchmarking can be defined formally as a continuous process of measuring products, services, and practices against a company's toughest competitors or companies renowned as company leaders, Camp (1992). The essence of benchmarking is captured effectively by the Japanese word *dantotsu* which means striving to be **the best of the best**. There are two generic terms of benchmarking. "Micro" benchmarks focus on specific processes. "Macro" benchmarks measure business characteristics and outputs that are the result of many practices and processes. Successful benchmarking strategies, however, need to link the two types of benchmarking under the general framework of diagnostic analysis.

The identification of units with exemplary performance will aid management to disseminate their operating practices to underperforming units. The assessment of market efficiency using DEA is based on DMUs that exhibit best observed practices, and therefore benchmarks would be identified among market efficient units. Market efficiency, however, is a necessary but not sufficient condition for a benchmark DMU.

A set of criteria can be defined in order to investigate the performance characteristics of individual market efficient units using as a basis the results of the DEA models solved for assessing the market efficiency of DMUs in M7.1 and M7.2 These criteria comprise issues related with the similarity between efficient and inefficient units and the relative distance of efficient units from the efficient frontier in their absence.

- **Similarity indices**

Market efficiency is assessed by M7.1 solving a linear programming problem for each DMU. In the solution process linear combinations of efficient units yield efficient targets that are compared with the observed performance of inefficient units (see session 5 in chapter 2). The frequency with which relatively efficient units are compared with inefficient units is an indicator of how *similar* is the input-output mix of individual efficient units with those of inefficient units. The more frequently an efficient unit is used as a comparator to inefficient units the higher is its input-output similarity and therefore can be trusted as a exemplary performer.

This similarity measure can be "weighted" further using the "*proportionate contribution*" of individual efficient units to the aggregate targets of those that are inefficient. This proportionate contribution is obtained using the values of the multipliers (λ's) from the solution to model M7.1. Let us denote by N_{AGR} the set of DMUs found efficient by M7.1. An efficient unit k contributes an amount of $\lambda_k^j y_{rk}$ to the target of output r of unit j where λ_k^j is the optimal value of the intensity variable of the efficient unit k when unit j is assessed. The aggregate proportionate contribution $C_{kr}^{j=1,\dots,n}$ of unit k to the targets of output r is given in (M7.3).

$$C_{kr}^{j=1,\dots,n} = \frac{\sum_{j=1}^{n} \lambda_k^j y_{rk}}{\sum_{k \in N_{AGR}} \sum_{j=1}^{n} \lambda_k^j y_{rk}} \qquad \forall k \in N_{AGR} \tag{M7.3}$$

Separate indices $C_{kr}^{j=1,\dots,n}$ are obtained for input-output variables used to assess market efficiency. Notice here that the magnitudes of the λ_k^j coefficients in M7.3 depend on the orientation (output expansion - input contraction) and/or the economic assumptions (constant - variable returns to scale) of the DEA model employed.

Estimation of $C_{rk}^{j=1,\dots,n}$ would rank efficient units on the basis of their proportionate contribution to the aggregate targets of individual inputs-outputs. These measures combined with the frequency that efficient units are used as peer comparators constitute *similarity* measures between efficient and inefficient units.

- **Relative closeness indices**

Another measure of distinction between relatively efficient units can be obtained by assessing some type of relative distance between relatively efficient units. This type of information is not provided by the aggregate and site-specific market efficiency models in M7.1 and M7.2 which give an efficiency of one for each efficient unit. This information, had it been available, it could be used to identify efficient units with exceptionally high distances from other efficient DMUs and hence rank efficient units on an "absolute" scale. This scale is derived as the minimum distance between relatively efficient units and the efficient frontier consisting of the remaining DEA efficient units and it will be called absolute ranking efficiency (ARE). The new type of performance comparison is obtained between the **current** performance of DEA-efficient units and their projection on the **adjusted** efficient frontier due to their absence from the production possibility set.

The assessment of efficient units' performance on their absence from the production possibility set has been addressed in previous research work by Andersen and Petersen (1993), Lovell *et al.* (1991) and Adolfson *et al.* (1991). The mathematical form of the "absolute ranking efficiency" (ARE) is given in M7.4 for the case of an output radial expansion model under constant returns to scale assumption.

Absolute ranking efficiency (ARE)	(M7.4)

$$\max_{\lambda_j, E_{j_o}} E_{j_o}$$

$$\sum_{\substack{j=1\\ j\neq j_o}}^{n} \lambda_j x_{ij} \leq x_{ij_o} \qquad \forall i$$

$$\sum_{\substack{j=1\\ j\neq j_o}}^{n} \lambda_j y_{rj} \geq E_{j_o} y_{rj_o} \qquad \forall r$$

$$\lambda_j \geq 0 \text{ and } E_{j_o} \textit{ free}.$$

The solution to M7.4 yields a radial expansion factor $E^*_{j_o}$ for outputs which can take values less, greater or equal to unity. This implies that the absolute ranking efficiency $ARE_{j_o} = 1/E^*_{j_o}$ will be less than unity for inefficient units as it would have been had the assessed unit j_O been included in the production possibility set. A value of $ARE_{j_o} = 1$ characterises units located on the efficient frontier without however, being extreme points of the production possibility set Values of $ARE_{j_o} > 1$ characterise DEA efficient units and indicate the extent to which individual efficient units "overperform" in comparison with other efficient units. As we shall see in the numerical illustration that follows the scores of $ARE_{j_o} > 1$ should not be interpreted blindly as measures of "strength" or "superefficiency" as they may indicate extraordinary operating practices.

The benchmarking criteria introduced above are illustrated next on the public houses of the case study.

4.3.1. Identifying benchmark pubs

Multi-unit organisations are organised in networks of activity centres and it becomes impractical for central management to focus on individual units' operating practices. The brewery used in this thesis manages a network of 2,000 pubs. The central management of the brewery were very keen to revise their accounting based methods currently used for identifying benchmark outlets. This was done utilising the results of the market efficiency analysis models.

The *aggregate* market efficiency model was selected in this study as the basis for obtaining benchmark pubs. Benchmark pubs were defined in view of their managerial and scale size performance and hence constitute role models from the central management's point of view. The superior performance of a benchmark pub reflects the appropriate selection of scale size of operation and also the recruitment of effective local management for the pub.

Table 7.1 contains aggregate market efficient pubs with a contribution to the total targets for turnover of at least 1%. Four benchmarking criteria are used to investigate the possible use of these pubs as the role models of the brewery. The "peer" frequency and the contribution of aggregate market efficient pubs to targets reflect the "*similarity*" indices between efficient and inefficient pubs. The higher the performance in these indices the more "similar" are the input-output characteristics of efficient pubs with the underperforming ones. The absolute ranking efficiency score (ARE) yields information on the "*closeness*" of individual efficient pubs from the efficient frontier of the production possibility set in their absence. Finally, the relative profitability (profit margin) of pubs provides useful information on the ability of pubs to convert their turnover into profit. The profitability criterion has complementary purpose as is a surrogate measure of the cost efficiency of individual outlets. This acknowledgement of the global character of performance of individual units (market-cost efficiency) can prove to be very important in identifying reliable benchmarks within retail networks.

Table 7.1
Criteria for identifying benchmark pubs

Pub	Absolute ranking efficiency*(%)	Contribution to targets (%)	Frequency as comparator	Profitability (%)
PUB78	139	37.00	88	37
PUB34	151	8.10	48	36
PUB52	109	5.70	18	26
PUB153	123	4.00	22	32
PUB22	126	3.80	20	33
PUB33	115	3.00	10	38
PUB134	129	2.61	18	38
PUB50	174	2.53	26	37
PUB121	116	2.43	15	30
PUB61	128	2.30	16	25
PUB67	108	2.30	19	19
PUB45	148	2.21	11	33
PUB19	113	2.12	11	31
PUB148	111	2.05	6	34
PUB4	175	1.65	13	0.05
PUB59	128	1.34	10	22

* Indicates the proportion of turnover excess produced by the corresponding pub as compared with the efficient frontier in its absence.

The performance of pubs on the four benchmarking criteria is variable with only frequency and target contribution being positively associated, having rank correlation of 0.76. An exception was the case of pub 50 which contributes disproportionately lower to the turnover targets of pubs for its frequency of appearing as "peer" pub. Upon further investigation it was found that pub 50 has a very extreme input mix as it has a very large size but operates under very adverse market conditions (very poor market potential).

Benchmark pubs show satisfactory profitability scores with pub 4 being a noticeable exception. The very high number of competitors for the market size of pub 4 is the main cause for being rated efficient. Pub 4 on the other hand does not generate enough sales to secure a satisfactory level of profitability.

Further insights concerning the performance of relatively efficient outlets is given by the absolute ranking efficiency indices. To illustrate this argument let us consider the case of pub 4 and pub 50 which have absolute ranking efficiency of 175% and 174%, and profitability of 0.05% and 37% respectively. The absolute ranking efficiency scores indicate that the two pubs have unusual characteristics. These characteristics, however, need to be investigated using complementary information. For example, pub 50, has more "similar" characteristics (e.g. market size, number of competitors, sales) with inefficient units and therefore is used more often as a peer comparator.

The illustration of the benchmarking criteria on the set of pubs shows clearly the complementarity between all four of them in diagnosing the operating characteristics of relatively efficient outlets. Relatively efficient units that show very poor performance on the benchmarking criteria should not be considered as good operating practices as their rating is based on the extraordinary conditions under which operate and not on their exemplary performance.

5. Conclusions

Chapter seven was the preamble in an attempt to develop frontier analysis-based decision support mechanisms in *for-profit* making MUOs. Decision support in *for-profit* MUOs has *diagnostic* and *planning* stages. Diagnostic analysis was the main theme of the chapter using as a steppingstone the concept of *market efficiency*, that is the ability of individual outlets to attract custom and generate sales. The market efficiency of DMUs is complemented by the *cost efficiency* which determines the success of converting turnover into profit. *Profitability measures* also give the profit profile of DMUs compounding elements of cost control and effective product pricing. The assessment of these concepts of performance was illustrated using a set of public houses from a large brewery in the UK.

Frontier analysis tools were used to assess market efficiency of different tier management within *for-profit* MUOs. This resulted in defining the *aggregate market efficiency* which reflects the ability to generate turnover and operate at the most productive scale size whilst *site-specific market efficiency* reflects the ability to generate turnover given the scale size of operation. Aggregate and site-specific market efficiency appeal to the central and local management of MUOs respectively and therefore convey valuable information concerning the diagnostic mechanisms of the performance of multi-level organisations.

The input-output sets used to assess the various components of market efficiency are based on factors representing the market conditions in the surrounding area of individual units and also factors representing the internal strength of these units to attract custom. A number of critical issues emanate from the development of input-output sets that could bias the efficiency measures obtained. These include the use and measurement of competition in the input/output set and the definition of appropriate market area of individual outlets.

Finally, the diagnostic analysis sought to provide an advanced framework for supporting benchmarking in retail MUOs. This framework combined information obtained from either the use of market efficiency models and/or from more dedicated analysis. Evidence from applying these techniques in identifying benchmark pubs showed that there are benefits from using more than one criterion for characterising exemplary performers.

The diagnostic based study on the performance of *for-profit* MUOs provide limited information concerning the development of decision support systems. The missing factor concerns the planning dimension of decision support which is the main theme of chapter eight.

- END OF CHAPTER SEVEN -

Chapter 8

Planning analysis and decision support for a-priori resource allocation[1]

1. Introduction

Chapter 8 focuses on planning problems in for-profit MUOs. The methods developed in the chapter seek to support decisions that require resource commitment (e.g. capital investment, reorganisation policies, etc.) and therefore information concerning their appropriateness is of vital importance to the management of MUOs.

The diagnostic analysis developed in chapter seven advocated the assessment of market efficiency as the main performance yardstick in for-profit DMUs. The *aggregate* and *site-specific* variants of market efficiency focused on ascertaining good operating practices and to estimate market efficiency targets for "underperforming" DMUs compatible with different tiers of management. This analysis, however, does not support the development of planning scenarios that would improve the performance of individual units by adjusting their operating profile. Elements such as the scale of operation of individual units, and the type/mix of services provided can affect adversely the units' long run viability. The way these issues are pursued in this chapter is discussed next in more detail.

---

1 Part of the models developed in this chapter are included in the paper "Performance improvement decision aid systems in retailing organisations using DEA", by A. Athanassopoulos, forthcoming in *Journal of Productivity Analysis.*

- **Improving the scale size of operation of DMUs**

The aggregate and site-specific market efficiency was assessed in chapter 7 as part of the diagnostic analysis of the operations of individual units. Constant and variable returns to scale economic assumptions were used as vehicles for disentangling the market efficiency of two tiers of management in multi-level MUOs. Aggregate and site-specific market efficiency seek to assess the extent to which decisions by different levels of management support individual DMUs to utilise their market potential.

The assessment of aggregate and site-specific market efficiency does not give any insights on whether individual outlets deploy their resources in proportions adequate to the market conditions under which they operate. The latter is an issue closely related to the concept of economies of scale as it corresponds to the extent that the scale size of controllable inputs can be improved to utilise the market conditions (uncontrollable inputs) of individual outlets.

The improved scale-size development proceeds adopting variable returns to scale economic assumption, as its main emphasis is on improving the outlets' scale size. It is assumed that individual outlets may underperform due to the mismatch between the scale of their controllable inputs and the market conditions in which they operate. The investigation examines the degree of scale size improvements for outlets already in operation. Thus the results of this type of analysis will be useful to the middle management of the organisation that is responsible for identifying/realising investment projects for improving the performance of individual outlets.

- **Assessing outlets' long run viability**

The viability of individual outlets is determined by their ability to contribute to the organisation's profits. Generation of profit, on the other hand, cannot be seen in isolation as it is compounded by elements of market and cost efficiency. Short and long run viability of outlets will be investigated using efficiency-profitability portfolio matrices. The classification of outlets into clusters with similar market efficiency and profitability will give the opportunity to create different viability scenarios for each cluster.

- **Assessing service mix effectiveness**

The range of services offered by retail organisations is largely determined by the market characteristics of individual outlets. Appraisal of the degree that different service mixes may lead into higher returns is always on the managerial agenda of service organisations.

Product mix differential cannot be appraised using sales increase information as it fails to capture other factors affecting performance. The assessment of market efficiency differential due to different service mix of individual units is advocated here as a more systematic way of assessing service mix effectiveness.

The remainder of this chapter seeks to address the three components of decision support introduced above. This is done by developing appropriate mathematical models which are illustrated using data from the public houses used earlier in chapter 7.

2. Assessing improved scale size (ISS) targets

The aggregate market efficiency model in chapter 7 compounds scale and managerial effects on the performance of units, whilst the site-specific market efficiency model eliminates these scale effects from market efficiency. The scale size in itself can be used, however, as a vehicle for exploring further improvements in the performance of inefficient units. This is based on the concept of the Most Productive Scale Size (MPSS) as introduced by Banker (1984), Banker and Morey (1986b) and generalised by Banker and Thrall (1992).

In chapter two it was shown that to each input/output production possibility (X,Y) corresponds at least one MPSS given its input/output mix. Estimation of the MPSS seeks to obtain the scale size that maximises the average productivity of the individual unit. **For instance, an input/output feasible combination (X_0,Y_0), under variable returns to scale assumption, is a MPSS if and only if <u>for all</u> feasible multiples of this combination (δX_0, εY_0) we have $\delta \geq \varepsilon$.**

Formulae M8.1 below can be used to estimate output expansion MPSS targets $\left(\breve{x}_{ij_o}, \breve{y}_{rj_o}\right)$ of inefficient outlets, Banker (1984).

$$\breve{x}_{ij_o} = \frac{x_{ij_o} - s_i^*}{\sum_{j=1}^{n} \lambda_j^*} \quad \forall i; \quad \breve{y}_{rj_o} = \frac{z^* y_{rj_o} + s_r^*}{\sum_{j=1}^{n} \lambda_j^*} \quad \forall r, \qquad \text{(M8.1)}$$

where $\left(x_{ij_o}, y_{rj_o}\right)$ is the level of the ith input and rth output of unit j_O and $z^*, \lambda_j^*, s_i^*, s_r^*$ are the optimal values of the *aggregate market efficiency* model developed in chapter seven and reproduced in M8.2 below.

$$z^* = \left\{ \max \ z \ \middle| \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{ij_o} \text{ and } \sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = z y_{rj_o}, \lambda_j \geq 0, z \textit{ free.} \right\} \qquad \text{(M8.2)}$$

Application of formulae M8.1 yields input/output targets beyond the target levels estimated by the *aggregate* market efficiency model which depend on the magnitude of the scale indicator $\sum \lambda_j^*$. This enables management to estimate the input/output mix that maximises the average productivity of each scale inefficient unit, Banker (1984). In other words, for each feasible input/output combination (δX_0, εY_0) the MPSS maximises the ratio ε/δ. **Targets assessed using formulae (M8.1) will be denoted as <u>improved scale size (ISS) targets</u> as they seek simultaneous adjustment to the input/output scale of inefficient units.**

The production technology employed by some units includes inputs/outputs that are not under direct managerial control. Banker and Morey (1986), Thanassoulis and Dyson (1992), and Thanassoulis *et al.* (1993), discuss this problem and recommend target setting models that do not violate the uncontrollable nature of some input/output variables. In the ISS in M8.1 the problem of input/output controllability is more obvious as inputs/outputs are adjusted *pro rata* by a factor of $1/\sum \lambda_j^*$. The scale adjustment of uncontrollable inputs/outputs needs to be prevented in order to give practically feasible planning targets.

A revised model for assessing ISS targets can be employed in M8.3 below, based on the assumption that inputs and outputs can be classified into controllable and uncontrollable groups. Let us denote I_c and I_f the controllable and uncontrollable subsets of the input set I=1,...,m respectively and x_{ij}^c and x_{ij}^f represent the i^{th} controllable (c) and fixed (f) input values of the j^{th} unit. Similarly, O_c and O_f are respectively controllable and fixed output subsets of the output set O=1,...,s with y_{rj}^c and y_{rj}^f denoting the r^{th} controllable (c) and exogenously fixed (f) output values of the j^{th} unit.

Table 8.1 lists, in the left column, the model M8.3 used to derive the ISS efficiency rating q^* and the scale size factor, $\sum_j \kappa_j^*$. In the right column of the same table are listed the formulae for deriving the ISS targets for individual units.

Table 8.1

Improved scale size targets with exogenously fixed inputs/outputs

(M8.3) ISS market efficiency with exogenously fixed inputs/outputs	Improved scale size (ISS) targets
$\underset{q,\kappa_j}{Max}\ q$ $S.t.\ \sum_{j=1}^{n} x_{ij}^{c}\kappa_j + w_i^c = x_{ij_o}^{c} \quad \forall i \in I_c$ $\sum_{j=1}^{n}\left(x_{ij}^{f} - x_{ij_o}^{f}\right)\kappa_j = 0 \quad \forall i \in I_f$ $\sum_{j=1}^{n} y_{rj}^{c}\kappa_j - w_r^c = q\, y_{rj_o}^{c} \quad \forall r \in O_c$ $\sum_{j=1}^{n}\left(y_{rj}^{f} - y_{rj_o}^{f}\right)\kappa_j = 0 \quad \forall r \in O_f$ $\kappa_j, w_i^c, w_r^c \geq 0$ and q free	$\breve{x}_{ij_o}^{c} = \sum_{j=1}^{n}\kappa_j^* x_{ij_o}^{c} \Big/ \sum_{j=1}^{n}\kappa_j^* \quad \forall i \in I_c$ $\breve{y}_{rj_o}^{c} = \sum_{j=1}^{n}\kappa_j^* y_{rj_o}^{c} \Big/ \sum_{j=1}^{n}\kappa_j^* \quad \forall r \in O_c$ $\breve{x}_{ij_o}^{f} = x_{ij_o}^{f} \quad \forall i \in I_f$ $\breve{y}_{rj_o}^{f} = y_{rj_o}^{f} \quad \forall r \in O_f$

The market efficiency is assessed by M8.3 having adjusted the uncontrollable inputs/outputs included in the study. This adjustment will enable the assessment of improved scale-size targets using the formulae listed in the right column of Table 8.1. The model in M8.3 is a generalisation of the Banker and Morey (1986) model and it differs from model M8.1 in the way exogenously fixed inputs and outputs are treated. This modification does not adjust the scale of uncontrollable inputs/outputs when ISS targets are estimated. **Improved scale size targets are estimated, therefore, for controllable[2] input/output attributes only.** The constraints that correspond to uncontrollable inputs/outputs are treated as equalities in order to avoid the presence of slack[3] adjustments. Thanassoulis *et al.* (1994) in a study of the provision of perinatal care in the UK reported positive slack values for uncontrollable

---

2 Controllability of input/output attributes depends on the hierarchy of organisational structure. For example, expansion of an outlet's capacity is not a controllable input for low level managers while it is under the discretion of higher level management.

3 The solution to model M8.3, however, includes a second phase optimisation as in the classic DEA models (see model M2.8 in chapter 2). In this second phase the slacks associated with controllable inputs/outputs are maximised by keeping the expansion factor q^* at the levels obtained in the first phase.

input/output variables unless modifications similar to the ones in M8.3 are applied (in Thanassoulis *et al.* (1994)).

Some general comments on the rationale and computational characteristics of the ISS model are provided below.

- *The Aggregate market efficiency and the Improved scale-size (ISS) are models with different purposes.*

Evidently the improved scale size model developed in M8.3 yields different efficient targets compared to those obtained by the aggregate market efficiency model in M8.2. A reinforcement of the purpose of the two models needs to be made in order to clarify their *raison d'être*. Model M8.2 assesses the aggregate market efficiency of units compounding managerial and scale efficiency. The targets estimated by M8.2 give the maximum feasible output expansion of inefficient outlets, compounding controllable and uncontrollable inputs and thus it shows how effective past decisions were to opening individual outlets at given locations. Model M8.3, adopts also a constant returns to scale economic assumption whilst it disentangles the effects of controllable and uncontrollable inputs on market efficiency. The differentiation between controllable and uncontrollable inputs/outputs will allow to assess performance targets that focus on improving the scale size of operation of controllable inputs/outputs (middle management responsibility).

- *The use of the ISS model requires modification of the production possibility set for each assessed DMU.*

The solution to the ISS model requires transformation of the original input/output mix of assessed units in order to preserve the status of uncontrollable inputs/outputs. The original input/output levels of, say, unit j_O are modified from $\left(x^c_{ij_o}, x^f_{ij_o}, y^c_{rj_o}, y^f_{rj_o}\right)$ to $\left(x^c_{ij_o}, 0, y^c_{rj_o}, 0\right)$ whilst on the other hand the inputs/outputs of observed units included in the production possibility set are modified from $\left(x^c_{ij}, x^f_{ij}, y^c_{rj}, y^f_{rj}\right)$ to $\left(x^c_{ij}, (x^f_{ij} - x^f_{ij_o}), y^c_{rj}, (y^f_{rj} - y^f_{rj_o})\right)$ which indicates that the production possibility set is customised for the assessment of each DMU j_O. This is **a unique characteristic of the ISS model as in all other "known" frontier analysis models (see chapter two), the inputs/outputs of the observed units of the production possibility set remain unadjusted during the assessment of individual units' performance**.

The rationale of the purpose-built production possibility sets used in M8.3 considers the performance of the assessed DMU j_O by eliminating the input/output levels $(x^f_{ij_o}, y^f_{rj_o})$ of uncontrollable inputs/outputs to a zero level. More importantly the uncontrollable inputs/outputs of the remaining observed units are modified according to the input/output levels of the assessed unit j_O. This customisation of the efficiency assessment seeks to alleviate the impact of the size of uncontrollable inputs/outputs on the performance of the assessed unit. The remaining units can be used as comparators only after an equivalent adjustment (by the levels of units j_O) has been made to their uncontrollable inputs/outputs. A more detailed investigation of the rationale of the ISS model can be found in the appendix of chapter 8 (appendix 8A).

2.1. Illustrating the improved scale-size model on real data

Decisions made in for-profit MUOs regarding the operation of business units are based on the assumption that their size and scale of operation should match the market conditions in their close vicinity. However, the rapid changes in the market conditions necessitate equivalent response from organisations in order to preserve/augment their market share in these markets. In the case of pubs, for example, a reduction/increase in the number of potential customers in its surrounding area would need to be responded to in terms of the size and service mix offered by the pub concerned. The improved scale-size targets can be used to develop a support mechanism for selecting a course of actions that would let units adapt to their changing marketplace.

The pubs of the brewery used in the previous chapter are used again to illustrate the method of the improved scale-size targets. The brewery invests annually a considerable amount of capital for altering the profile and style of its outlets and thus assessment of improved scale size targets may provide considerable support to the capital allocation process.

The input/output set used to obtain results from M8.3 is the one developed originally in chapter 7 (see Figure 7.6). M8.3, however, requires the separation of the input factors on the basis of their long run controllability which is described in Table 8.2.

Table 8.2

Controllable & uncontrollable inputs for ISS targets

Controllable Inputs (x^c_{ij}, i=1,2,3 in model M8.3)	**Uncontrollable Inputs** (x^f_{ij}, i=1,2,3,4 in model M8.3)
• Bar area (ft^2)	• Consumption of alcohol in the surrounding area (barrels)
• State of repair of the pub	• No. of potential customers
• No. of car park facilities	• Gross income of households in the surrounding area (£)
	• Competition

The solution of model M8.3 adds a new set of targets to the current aggregate and site-specific market efficiency targets for each operating unit. As the scale size of operation of individual units is determined by central management, the discussion of the ISS results will be compared with the corresponding aggregate market efficiency targets estimated in chapter seven (see M7.1). Two particular pubs have been selected to illustrate the model used and the differences in target estimates obtained by the two models. Summary results will also be given focusing on average statistics of the targets of the sample of 154 pubs.

- *The improved scale-size (ISS) targets do not always yield pro-rata scale size expansion/reduction to inputs/outputs.*

Expansion/contraction of the input/output mix using the ISS formulae (defined in Table 8.1) is estimated using the optimal expansion factor q*, the returns to scale indicator, $\sum \kappa^*_j$ and the potential slack variables in the optimal solution of model M8.3. The importance of the slack values in estimating planning targets is illustrated in Table 8.3 using pub 15 as an example.

Table 8.3

Improved scale-size targets for pub 15 ($\sum \kappa^*_j$=0.75, q*=1.3)

Variable	**Current**	**Improved scale-size targets** **Current**	**Slack**	**Σκ**		**Target**
Bar area (ft^2)	1257	(1257 -	440)	/ 0.75	=	***1088***
Car park facilities	38	(38 -	8.4)	/ 0.75	=	*39*
State of repair	3	(3 -	0)	/ 0.75	=	*4*
Turnover (£)	97790	((97790 +	0) * 1.3)	/ 0.75	=	*170158*

The estimated ISS targets of pub 15 provide the input/output mix that would render pub 15 at its most productive scale size. Pub 15 is scale inefficient and it should expand its scale size (i.e. $\sum \kappa_j^* < 1$). One would expect, therefore, that a proportionate expansion of its inputs would result in a more than proportionate increase to its output. For the bar area of pub 15, however, this is not the case. The positive slack value (440 ft^2) exceeds the impacts of the scale factor, $\sum \kappa_j^*$ and thus, pub 15 should **reduce** its bar area and **increase** the scale of all other controllable inputs in order to maximise its average productivity.

From the illustration above one can conclude that the scale indicator, $\sum \kappa_j^*$, is not a sufficient proof indicator of the direction of the scale-size adjustments estimated by the ISS model.

- *The scale-size adjustments may result in reduced output levels for inefficient units.*

Improved scale-size targets for inefficient units could, some times, recommend reductions to the levels of outputs, in contrast to the targets traditionally estimated by models like M8.2. This can be illustrated using the case of pub 131, exhibited in Table 8.4.

Table 8.4

Aggregate & ISS targets for pub 131

Variable	Current	**Aggregate** targets ($\sum \lambda^*$ =0.78, z_o=1.21)	**ISS** targets ($\sum \kappa^*$ =1.19, q_o=1.054)
Bar area (ft^2)	1003	1003	841
Car park facilities	16	11	13
State of repair	11	8	9
Turnover (£)	187851	227001	166382

The estimated targets for pub 131 differ as a result of the different assumptions made by models M8.2 and M8.3 respectively. The aggregate targets sought to expand the turnover of the assessed pub without adjusting its current scale size (the aggregate targets are estimated using the composite unit $\hat{x}_{ij} = \sum \lambda_j^* x_{ij}$, $\hat{y}_{rj} = \sum \lambda_j^* y_{rj}$ from M8.2). On the other hand the ISS targets sought to adjust the scale size of operation of the inefficient unit. The different purpose of the two models is reflected in their estimated targets.

Pub 131 has excess size capacity (bar area) for its market size and one needs to explore the prospects of its future market size before any actions are taken. The reduced turnover levels, as a consequence of the 20% reduction on its bar area, can be justified so long as the controllable costs will be reduced to a higher level and thus the alterations will increase profitability. This information, however, cannot be provided by the market efficiency models and therefore the relation between outlets' size and profitability needs to be explored separately.

2.1.1. *Summarising the improved scale-size results for the set of 154 pubs*

Having discussed the mechanism of the ISS model we can give some summary statistics of the targets estimated for the set of 154 pubs in our example. Table 8.5 shows results of the average estimated targets for the turnover and the controllable inputs I_C.

Table 8.5

Aggregate & Improved scale size targets (average)

	Recommended planning policy over input mix								
	Contraction $\left(\sum \kappa_j^* > 1\right)$			*Expansion* $\left(\sum \kappa_j^* < 1\right)$			*No change* $\left(\sum \kappa_j^* = 1\right)$		
Input/output	**Current**	**Aggregate**	**ISS**	**Current**	**Aggregate**	**ISS**	**Current**	**Aggregate**	**ISS**
Bar area (ft^2)	1572	1295	1226	980	872	1114	1190	1122	1190
Car parks	29	26	18	17	12	16	20	19	20
State of repair	11.4	10	9.1	10	9	12	10.5	10	10.5
Turnover (£)	183740	277847	228650	144960	217480	244454	187970	199772	187970
Efficiency		63%	67%		67%	74%		93%	100%
No. of pubs		55			52			47	

The indicator of scale adjustments, $\sum \kappa_j^*$, is used as a criterion for classifying the pubs in Table 8.3. The column headed "Current" corresponds to the average observed input/output values of each class of pubs. The columns headed "aggregate" and "ISS" correspond to the average target values estimated by models M8.2 and M8.3 respectively. About 69% (107 out-of 154) of the pubs have ISS targets that recommend scale adjustments.

A general pattern emerges from Table 8.5 regarding the pubs' size and the associated targets for the three classes of pubs. Pubs in the contraction cluster are relatively large pubs in all controllable input dimensions. This argument is consistent with an earlier finding in chapter seven where a negative statistical association between the aggregate market efficiency of pubs and their bar area was found (see Appendix 7B).

Targets obtained from the ISS model concerning the state of repair of pubs must be interpreted with caution. The nature of this variable is twofold. State of repair is an internal input reflecting the attractiveness of individual pubs. Achievement of a level of state of repair requires consumption of resources such as capital investment and maintenance. The state of repair of pubs has a decreasing life span[4], and therefore, its targets can be used to obtain either minimum standards of state of repair before action is taken or estimate the necessary improvements for attracting more custom.

2.1.2. *Practical use of different market efficiency models*

The assessment of market efficiency targets in chapters seven and eight is made using three alternative models. These models, namely aggregate, site-specific and improved scale-size have complementary roles in supporting managerial diagnosis and planning in MUOs. Profit making MUOs have in practice multi-level managerial structure and thereby the assessment of their market efficiency should be customised for different levels of management.

The aggregate and site-specific market efficiency models can distinguish market efficiency between central and local management in a diagnostic role. A third market efficiency component, namely ISS, seeks to explore performance improvements by appropriate alterations to the scale size of inefficient units. These decisions are mainly left to the middle management of MUOs which are responsible for implementing the strategic plants adopted by central management.

For the management of the MUO, however, the three DEA models give a unique opportunity for decomposing the performance characteristics of individual units to assess different tiers of management. This would enable management to detect sources and causes of weak and/or superior performance, and therefore to manage more effectively its network of outlets.

---

4 For example, the outlets' decor wears off over time due to consumers changing behaviour whilst the furniture needs replacement in regular time intervals.

2.2. Market efficiency & returns to scale

Evidence concerning the presence of returns to scale in the operation of pubs gave conflicting results depending on the DEA model adopted for the analysis. These results, however, provide very useful insights concerning the causes of scale inefficiencies of pubs. Chapter two gives technical details (see Table 2.1) on how the presence of returns to scale can be investigated based on Banker *et al.* (1984) and Banker and Thrall (1992) criteria.

In the pubs' case study the aggregate market efficiency model, M8.2, yields scale efficiency estimates on the basis of **all** input factors. Thus increasing or decreasing returns to scale can prevail due to non-optimal mix of controllable (i.e. bar area) and uncontrollable (e.g. consumption of alcohol in trading area) input factors. Results obtained by M8.2 provide very useful information concerning the effectiveness of decisions made by central management in the past to open and locate outlets (aggregate market efficiency).

If a mismatch is diagnosed, by model M8.2, between the outlet's profile and its surrounding area, then the improved scale-size model (M8.3) is appropriate for deriving the necessary and feasible scale adjustments of controllable inputs in accord to the characteristics of its market. Thus, the scale size targets derived by M8.3 focus on the optimal scale of controllable inputs given the scale of those that are uncontrollable.

Table 8.6 summarises returns to scale results as obtained by the aggregate market efficiency and ISS models respectively.

Table 8.6

Estimating returns to scale of pubs

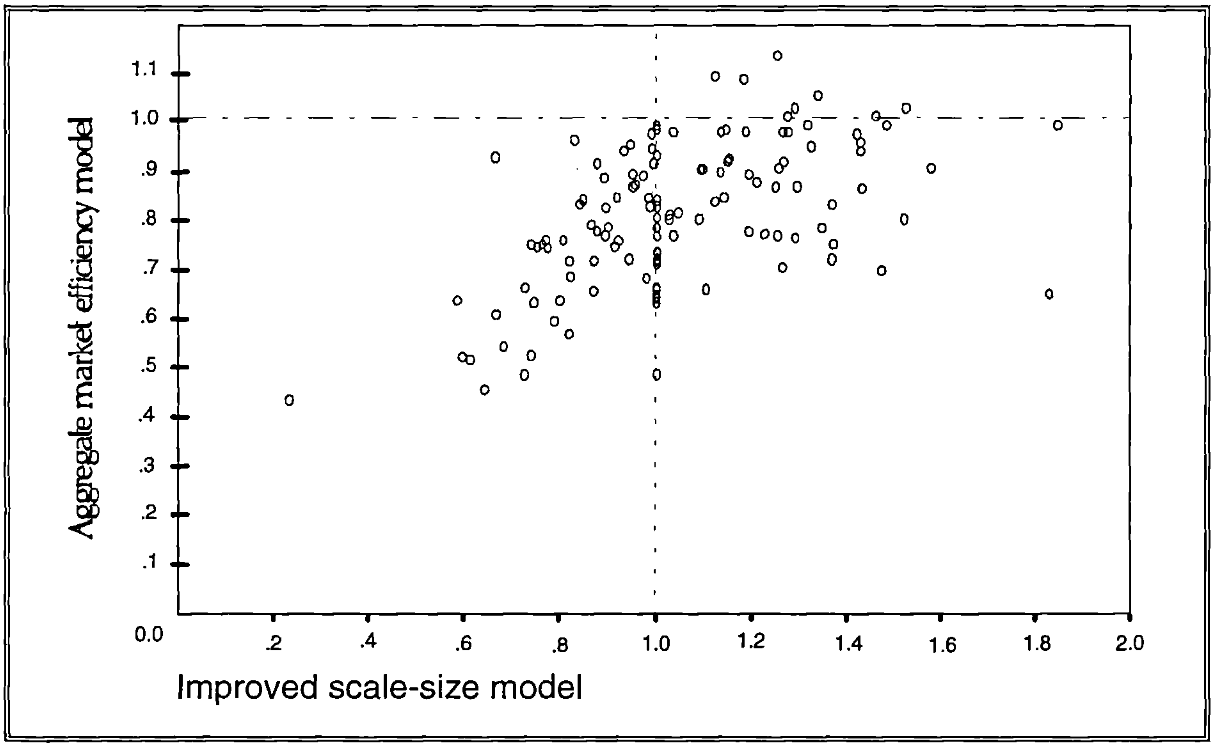
		Aggregate market efficiency $\sum\lambda_j^*$			
		Increasing	Decreasing	Constant	Total
Improved Scale Size $\sum\kappa_j^*$	Increasing	52	0	0	52
	Decreasing	45	8	2	55
	Constant	15	0	32	47
	Total	112	8	34	154

Table 8.6 reveals the presence of scale inefficiencies in the operation of public houses. For many pubs, however, the results obtained by the two models are vastly different. Insofar as

the aggregate market efficiency returns to scale are concerned, one would argue that relatively many pubs operate under local increasing returns to scale. This would give an indication to central management for expanding the scale size of units. This scale indication, however, compounds controllable and uncontrollable characteristics of individual pubs which are beyond the discretion of management.

The ISS model yields considerably different estimates of returns to scale; only 52 out of the previously 112 scale inefficient pubs were found operating under increasing returns to scale. The removal of the uncontrollable input factors' bias, therefore, has had considerable effect on the characterisation of returns to scale. Additional information on the relation between the returns to scale obtained by the two models can be found in Figure 8.1.

Figure 8.1
Aggregate ($\sum \lambda_j^*$, M8.2) versus ISS ($\sum \kappa_j^*$, in M8.3) scale factors



The bivariate plot in Figure 8.1 shows how the returns to scale factors obtained by models M8.2 and M8.3 are associated. The two returns to scale factors did not have a strong linear association (Pearson and Spearman rank correlation coefficients were found 0.41 and 0.43 respectively). There is a linear pattern between the scale factors of pubs operating under increasing returns to scale in both models (see lower-left quadrant). This pattern is different for a considerable number of pubs which were assessed operating under increasing returns to scale by M8.2 but decreasing returns to scale by the M8.3 model (right-down quadrant).

3. Assessing outlet viability

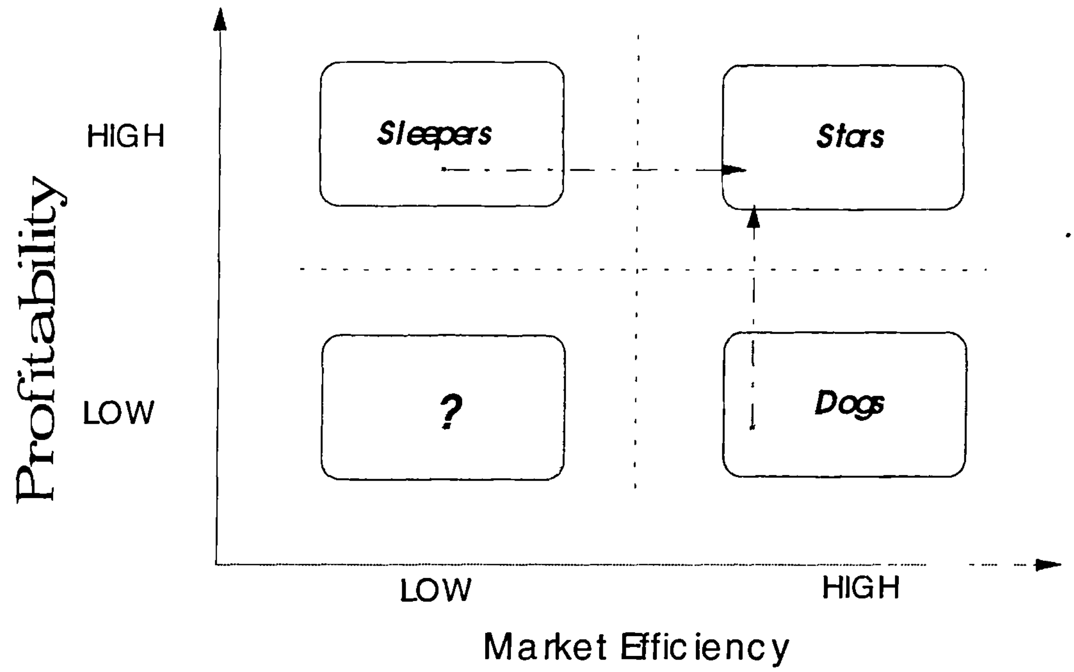
The commercial viability of individual outlets determines their ability to support the global organisational objectives. Their commercial viability is determined by the individual characteristics of individual outlets and more generally by the effectiveness of the services/products marketed by the entire organisation. The investigation at the individual unit level will be made using product-portfolio matrices and also some models for "predicting" the profit implications of changing the input/output mix of inefficient outlets in line with the ISS target estimated earlier on.

As the size of outlets in MUOs can be excessive a taxonomy for classifying the network of outlets into clusters with similar viability prospects is necessary. In the context of the present study market efficiency and profitability are used as criteria for assessing units' viability using two dimension portfolio matrices.

This is demonstrated in Figure 8.2 where operating outlets are classified into four clusters based on their market efficiency and profitability. Each cluster represents outlets with different operating profiles and thus different viability prospects.

Figure 8.2

Market efficiency Vs. Profitability



Market efficiency is used in Figure 8.2 to reflect the extent to which individual outlets realise their market potential by generating turnover. Profitability on the other hand is a relative measure (as opposed to gross profit which is an absolute measure) which focuses on the outlets' relative ability to convert generated turnover into profit (cost efficiency).

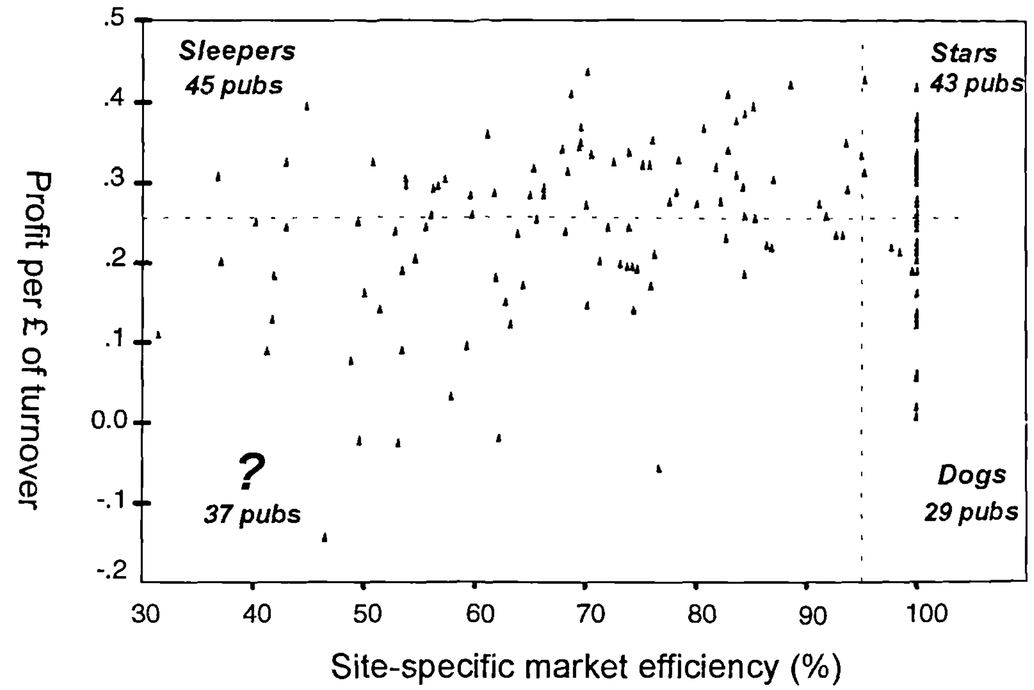
A methodological problem arises from the selection of the appropriate model to represent market efficiency. Profitability reflects an ability to control costs and, therefore, it is a measure with primary appeal to local management (given the scale size of the unit). Therefore, assessment of profitability against site-specific market efficiency focuses on outlets' viability from a local management's point of view. The likelihood of improving performance relies on units' ability to improve local management and control costs. We shall indicate this as the ***short run viability*** of individual units.

On the other hand profitability against a measure of "aggregate" market efficiency brings scale size factors into the assessment of viability. This assessment appreciates the importance of central management decisions on the outlets' long run viability. The improved scale-size efficiency model in, M8.3, should be chosen to reflect the efficiency of units as it preserves the uncontrollable nature of some input factors included in the analysis. This type of assessment concentrates on the ***long run viability*** of individual outlets as the potential performance improvements would require capital investment decisions which have a long time horizon of implementation.

Some illustrative results concerning the assessment of pub' viability are shown in Figure 8.3 (short run viability) and Figure 8.4 (long run viability). Pubs are classified into four viability quadrants defined in Figure 8.2. The median profitability of pubs (0.25) was selected as an arbitrary cut-off point for defining the upper and lower profitability clusters.

Figure 8.3

Short-run viability: Site-specific market efficiency Vs. profitability



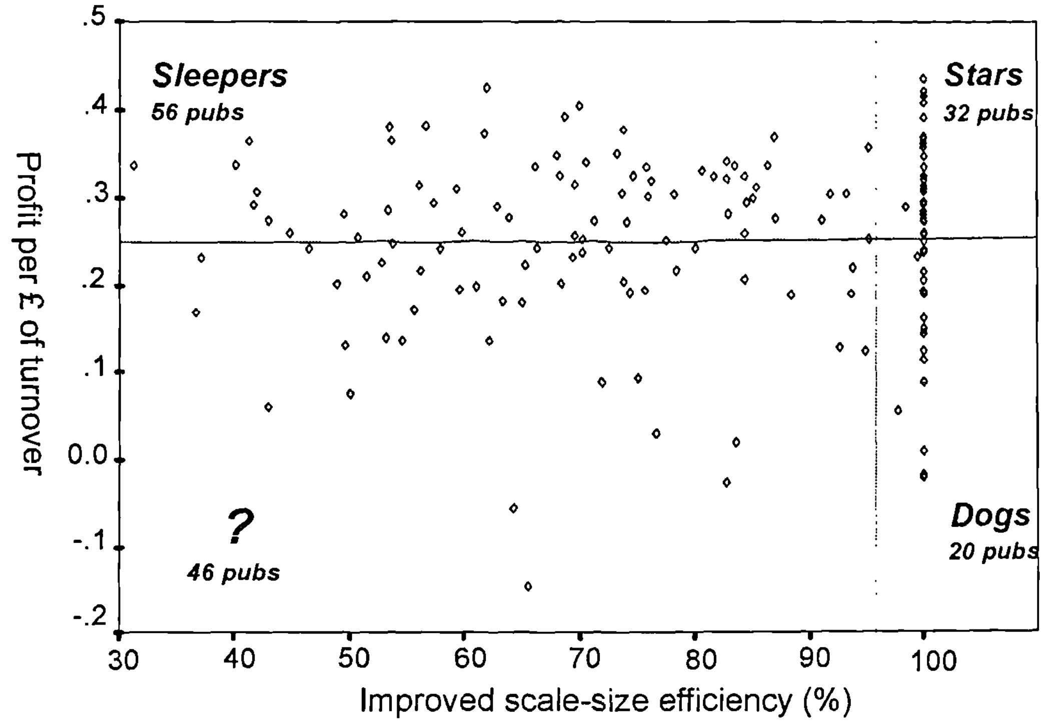
Different policy implications apply for pubs found within each of the four viability-quadrants in Figure 8.3. The weak positive linear relationship between profitability and site-specific market efficiency supports the claim that local management of profit making outlets can influence their profitability only marginally.

Low site-specific market efficiency and profitability characterise underperforming outlets (pubs) in both performance dimensions. On the other hand, high efficiency and low profitability indicate either poor cost control or inadequate scale size of operation (large outlet's size in an area with adverse market conditions). High profitability and low efficiency identify outlets that need to increase market share.

Further analysis on the *long run viability* of outlets of can be obtained using the ISS estimates of market efficiency. This information is provided in Figure 8.4.

Figure 8.4

Long run viability: Improved scale-size efficiency Vs. profitability



Extending the efficiency assessment of pubs towards including controllable scale factors increases the dimensions taken into account in the viability assessment. Simple comparison of Figure 8.3 and 8.4 concerning the distribution of pubs in the four quadrants shows that twenty pubs change cluster membership depending on the model employed for assessing efficiency. These pubs are subject to scale size adjustments for improving their viability.

Although the evidence provided in Figures 8.3 and 8.4 is an illustration of a particular set of data, the general message is that the assessment of outlets' viability needs to take into account alternative definitions of efficiency. The product portfolio matrix alluded to above gives preliminary signals on the relative positioning of individual outlets, and therefore the general status of the MUO. A set of alternative scenarios need to be developed for groups of outlets on the product portfolio matrix. These scenarios will seek to address performance issues of the network taking into account corporate and strategic issues.

3.1. Profits impacts on scale-size strategy (PRISS)

The development of methods to provide adequate decision support in profit making organisations has already been mentioned in chapter one using corporate databases like PIMS. These studies seek to obtain regression based forecasts of firm profits as a function of various (quantitative and qualitative) explanatory factors. Similar problems exist in the interior of MUOs insofar as the profits of individual outlets are concerned. A regression based methodology can also be used for assessing the impacts of the ISS targets on the profits generated by individual outlets.

A combined process, named profits impacts on scale-size strategy (PRISS), can be used to estimate the impacts of planning decisions on the current profits generated by individual DMUs. The rationale of the PRISS process is inextricably linked with the assessment of scale-size targets for inefficient outlets. The realisation of scale-size adjustments requires resource commitments in the form of capital investment projects that need to be evaluated using investment appraisal methods. Knowledge on the expected profits of alternative scale adjustment projects would complement the evaluation of corresponding capital investments.

A regression based method for assessing the expected profit returns, $E(P_j)$, as a result of the scale-adjustment policies, can be developed as follows:

$$E(P_j) = c + \sum_{i \in I_c} a_i \breve{x}_{ij} + \sum_{r \in O_c} b_r \breve{y}_{rj} + \sum_c \gamma_c z_{cj} + v_j$$

$E(P_j)$	expected profit of unit j
c	constant term
$\breve{x}_{ij}$	is the difference between the observed x_{ij} and the ISS target for input i $$\breve{x}_{ij} = x_{ij} - \left(\sum_j \kappa^*_j x_{ij} \Big/ \sum_j \kappa^*_j \right),$$
$\breve{y}_{rj}$	is the difference between the observed y_{rj} and the ISS target for output r $$\breve{y}_{rj} = y_{rj} - \left(\sum_j \kappa^*_j y_{rj} \Big/ \sum_j \kappa^*_j \right)$$
z_{cj}	Independent variable *c* affecting the profit of unit j
a_i, b_r, γ_c	Regression coefficients for i^{th}, r^{th} and c^{th} variables respectively,
v_j	Random noise error term

The regression model above includes a statistical error term $v_j = N(0, \sigma^2)$. More advanced assumptions can be made concerning either the functional form of the regression model or decomposing the error term into random noise and efficiency components, see Lovell (1993).

The rationale of the PRISS method emanates from the strategic planning literature and the use of the PIMS data base. It is also affected by the relatively new trend within DEA for expressing the efficiency of assessed units as a function of various explanatory factors. Lovell *et al.* (1991) sought to link schools performance with environmental and socio-economic factors, Cooper and Galliegos (1992) sought to explain the variation of the assessed performance targets of airline firms as a function of input factors, Elyasiani *et al.* (1994) sought to identify the association between assessed efficiencies and multidimensional ratio analysis in the banking industry. The methods chiefly used for this analysis are alternative specifications of regression analysis (e.g. Ordinary least squares and Probit and Logit regression models).

PRISS deviates from the applications reported above in that the

> current methodologies seek to explain the efficiency of units as a function of various factors. **PRISS, on the contrary, seeks to identify the impacts of the targets assessed by DEA on dependent variables (e.g. profit) not included in the DEA assessment.**

The PRISS regression model is illustrated next on the pubs' data set by setting up a profit function. It should be noted that the profit function does not include prices of products or labour as the pricing policy is monitored by central management. Results obtained from the regression analysis model are summarised in Table 8.7 below.

Table 8.7
Expected profits from ISS targets

Groups of variables	Predictors	Expected gross profit (£)	T-statistic
Market conditions	Consumption of alcohol per competitor	0.002 **	2.65
	Potential customers	0.55 **	2.02
	Average income in surrounding area (£)	2.3 **	2.69
Internal characteristics of pubs	Observed Turnover per Labour costs	38123 **	14.55
	Difference in turnover (£)	0.09 **	2.66
	Difference in Bar area (ft^2)	4.25 **	4.30
	Difference in State of repair	- 1393 *	- 1.65
	Service mix {0,1} (Food service = 1)	9642 **	3.5
	Difference in car park spaces	207 **	2.8
	Quality of pub manager ([1, 30] scale)	2042 **	3.9
Alternative locations of pubs	Back street {0,1}	- 3793 *	- 1.65
	Suburban {0,1}	- 5413 **	- 3.40
	Industrial Estate {0,1}	- 3600 *	- 1.66
	Constant term	-120753 **	- 9.98
	R^2-adj.	**79 %**	

* Significant level of less than 10%

** Significant level of less than 5%

The regression model in Table 8.7 includes three families of explanatory variables: market conditions, adjustments in the controllable inputs/outputs of pubs, and spatial impacts due to alternative locations. The model has a satisfactory fit explaining up to 79% of the variation of current profits. No multicollinearity problems were detected[5] and therefore an interpretation of the partial regression coefficients can be made. The very large negative constant term (£-120,753) reflects to some extent the average level of generated turnover that is necessary to cover the average fixed costs.

---

[5] Variance inflation factor (VIF) <<< 3.

The **positive** association between profit and market conditions and between profit and the positive alterations on the size of internal characteristics of pubs was expected. The negative sign of the state of repair, however, reflects the cost-impacts of improving the state of repair of pubs in the short run. The difference between the observed and the ISS estimated turnover has a positive impact of 0.09 on the generation of profit which implies that each £1000 of extra difference in turnover would generate on average £90 of extra profit, all else being equal in the PRISS equation.

The PRISS model can also be used to "predict[6]" the expected profits of individual DMUs, using the regression coefficients listed in Table 8.7. The estimated expected profits for different classes of outlets are summarised in Table 8.8.

Table 8.8

Improved scale-size & profit implications (Overall sample set)

Planning recommendations	Current average profits (£)	Expected average profits (£)	Aggregate change (%)	Number of pubs
No change $\sum_j \kappa_j^* = 1$	52646	55216	+ 4.6	47
Expansion $\sum_j \kappa_j^* < 1$	35309	41225	+ 14.3	52
Contraction $\sum_j \kappa_j^* > 1$	51949	53980	+ 3.7	55

Results are provided in Table 8.8 in three different groups. Each group, namely no change, expansion, contraction, corresponds to a particular scale adjustment policy suggested by the ISS model. The estimated profit figures were obtained as follows. For the 47 pubs that operate at an optimal scale (no planning targets were estimated) the expected profit was truncated to be: **the observed profit**, for cases that the regression model had anticipated to be less than the observed profits **or the estimated profit by regression**, for cases where

6 An attempt was made to split the residual term v_j into inefficiency (u_j) and random noise (e_j) terms. The regression model, however, rejected the hypothesis of skewness of the inefficiency term (u_j) and therefore the statistical evidence allow only the presence of random noise within the current data set. The latter, seems to reinforce a managerial belief within the brewing industry that market (and not cost) efficiency is the main source of inefficiency of public houses.

the observed profit was smaller than the regression estimate. This adjustment was necessary as no scale adjustments were suggested for these pubs. For the other two categories of pubs the reported estimated profits are as obtained by the regression model.

For pubs with recommended expansion to their scale size there are positive profit increases by 14.3%. For pubs with contraction projects, (No. 55), or no planning projects (No. 47) a moderate profit increase of 3.7% and 4.6% is expected respectively. This implies, however, that even market efficient pubs should endeavour augmentation in the generation of gross profit.

Overall, the selection of regression analysis for developing the PRISS system could be criticised for estimating average and not efficient levels of expected profit. Regression analysis, however, is seen as more appropriate as the scale size adjustments included in the analysis do not eliminate all possible sources of inefficiency (e.g. site-specific) of individual units. Moreover, regression analysis is more flexible in dealing with qualitative type of information and also avoiding a-priori assumptions on the direction of association between profit and explanatory factors. In summary, the regression based approach of PRISS is seen to be appropriate in identifying patterns of association and estimating expected profit returns by allowing inefficiency in the operation of individual outlets.

4. Assessing service-mix effectiveness

Market efficiency has been advocated as a key measure of performance of profit making MUOs. The various components of market efficiency (e.g. site-specific) were defined, hitherto, at the DMU level. Thus, issues related with target setting, benchmarking, and long run viability were addressed focusing on individual DMUs. Another important aspect of performance, namely service mix effectiveness, is examined next which concerns the appropriateness of the service portfolio of the organisation.

> *The Service-mix effectiveness of individual units seeks to assess the extent to which there is market efficiency differential between units with a different mix of services.*

In the grocery industry for example service-mix effectiveness relates to the extent to which groceries with increased non-food services have market efficiency differential. In the brewing industry, on the other hand, service-mix effectiveness of pubs relates to the extent to which public houses providing food and drink services have market efficiency differential over those that focus mainly on the provision of drink services.

One could argue that the question is trivial in the sense that units with different service mix should be assessed separately and thus their market efficiency differential would emanate by the separate assessments. An investigation on a real life organisation (public houses) is made next to show that the assessment of service-mix effectiveness is not a straightforward problem as it was found that separating the units into groups of different service mix does not give any insights into their service mix differential.

The provision of food services by pubs is currently an important trend within the brewing industry. The actual impact of food sales on the performance of public houses has, however, yet to be assessed. The regression model used in chapter 7 (see Appendix 7A) found pubs with food services had on average an extra £42,147 of annual turnover. This, however, does not provide conclusive evidence on the actual performance of these pubs and more in depth investigation needs to be carried out. Heretofore, all market efficiency DEA models were based on a common production possibility set and thereby efficient frontier respectively. Table 8.9 below shows the extent to which these results are different for units with different service mix.

Table 8.9

Service-mix impacts on market efficiency

Market efficiency	Average (%) efficiency of pubs with food sales	Average (%) efficiency of pubs with no food sales	Welch's test t-statistic (p-value)
Aggregate	76.3	73.8	-0.78 (0.438)
Site-specific	84.6	83.6	-0.34 (0.731)
Improved scale-size	82.1	75.7	-2.01 (0.04)
No. of pubs	*89*	*65*	

Statistical evidence provided in Table 8.9 shows that the different service mix, notably alcohol and food, affect only the ISS efficiency (model M8.3). Insofar as the aggregate and site-specific market efficiencies are concerned there is no statistical evidence to show pubs' performance differs with service mix. Overall there is no conclusive evidence on whether there is market efficiency differential between outlets with different service-mix. A more thorough investigation will be employed in the next section to pursue this question more systematically.

4.1. Non-parametric models for service-mix effectiveness

Service mix-effectiveness will be assessed using the methodology of program efficiency introduced by Charnes *et al.* (1981) and developed further by Grosskopf and Yaisawarng (1990), and Kittelsen and Forsund (1992). These are DEA related methods with the main objective of comparing the efficient frontiers of groups of units with different service-mix.

Service-mix effectiveness is assessed using a two-stage process described below. In the first stage DMUs are classified into groups with similar service-mix; given a set of units N we create t=1,...,T subsets of units $N \equiv N_1 \cup ... \cup N_T$. The market efficiency $E_{j_o}^{N_t} = 1/z^*$ of unit j_O is assessed against a frontier made of units lying in the same group (N_t) using M8.4.

$$z^* = \left\{ \max z \left| \sum_{j \in N_t} \lambda_j x_{ij} \le x_{ij_o} and \sum_{j \in N_t} \lambda_j y_{rj} \ge z y_{rj_o}, \lambda_j \ge 0, z free \right. \right\} \qquad \textbf{(M8.4)}$$

In the second stage the efficiency of all units is re-assessed in a joint group where the inputs/outputs of individual units, say j_O, are previously adjusted to the target levels $\left(x_{ij}^a = \sum_{j \in N_t} \lambda_j^{*j_o} x_{ij}, y_{ij}^a = \sum_{j \in N_t} \lambda_j^{*j_o} y_{rj} \right)$ estimated by M8.4, where $\lambda_j^{*j_o}$ is the optimal multiplier used to obtain the contribution of unit j to the targets of the assessed unit j_O. The revised efficiency of unit j_O, $\tilde{E}_{j_o}^{N_t} = 1/\tilde{z}^*$, can be obtained solving the model in M8.5.

$$\tilde{z}^* = \left\{ \max \tilde{z} \left| \sum_{j \in N} p_j x_{ij} \le x_{ij}^a \ and \sum_{j \in N} p_j y_{rj} \ge \tilde{z} y_{rj}^a, p_j \ge 0, \tilde{z} free \right. \right\} \qquad \textbf{(M8.5)}$$

where $\left(x_{ij}^a, y_{rj}^a \right)$ are the adjusted levels of input i and output r using the solution of model M8.4. **The efficiency index estimated from M8.5 gives an estimate of the comparison of the efficient frontiers of units within the same group of service mix**. Inefficiencies estimated by model M8.5 are attributed to the potential superiority of different types of service-mix. On the methodological side the *service-mix effectiveness* models in M8.4 and M8.5 compound local and scale performance characteristics by adopting constant returns to scale assumptions. It was felt that the type of performance assessed in M8.4 should include all possible factors that affect service-mix effectiveness. Investigation of scale differences between the efficient frontiers of groups with different service mix can also be employed in order to decompose the performance differences between groups of outlets.

The results provided in Table 8.9 compared the market efficiency of outlets with different service-mix obtained from a single data set without separating them into clusters of similarity. A more in-depth investigation can be made using the models M8.4 and M8.5. Thus each pub in the study will be attached three different indices of aggregate market efficiency: one based on the aggregate set of units, one within sets of similar service mix and, finally, one on the adjusted aggregate set. Results of the three average aggregate market efficiencies are presented in Table 8.10.

Table 8.10

Assessing service mix effectiveness

Average **aggregate** market efficiency (*St.dev*)	Pubs with food sales (Sample size 65)	Pubs with no food sales (Sample size 89)
All pubs in the assessment[i]	76.3 (*21.3*)	73.8 (*22.6*)
Efficiency of pubs with similar type of services[ii] (M8.4)	81.1 (*17.1*)	82.3 (*16.9*)
Assessed pubs are efficient within their group of service mix[iii] (M8.5)	99.5 (*2.5*)	87.6 (*10.2*)

i : Efficiencies are estimated from the total sample of pubs (no. 154)
ii: Efficiencies are estimated independently between pubs with and without food sales
iii: Efficiencies are estimated from the total sample of pubs with inputs/outputs the targets estimated from phase (ii).

The results in Table 8.10 illustrate the importance of the tests suggested earlier on for assessing service-mix effectiveness. For example results from the first two models i.e. when all pubs were assessed together or separately using their row data, indicate that there is no market efficiency differential due to different service-mix.

Assessment of market efficiency separating pubs into groups with and without food services, yield expected higher efficiencies, but without substantial differences between the two groups of pubs. The increased efficiency in both samples indicates that in the joint assessment efficient pubs are used as "peers" of inefficient pubs with different service mix (e.g. inefficient pubs with food services are outperformed by efficient pubs without food services and *vice versa*).

Results in the third row were derived by M8.5 where pubs were assessed jointly with adjusted inputs/outputs at the target levels obtained by M8.4. The market efficiency of the two groups is substantially different. The efficient frontier of pubs with food sales outperforms systematically the corresponding efficient frontier of pubs without food sales. The differential market efficiency due to different service-mix did not arise in the first row of Table 8.10 as many pubs with food services where inefficient. When these inefficiencies were eliminated using model M8.4 the revised assessment by model M8.5 showed clearly the performance strength of pubs providing food services. **This potential would have remained hidden had the service-mix effectiveness study not being employed.**

5. Conclusions

The main objective of this chapter was to complement the diagnostic analysis employed in chapter seven for supporting planning decisions in for-profit MUOs. These decisions are primarily concerned with the adjustment of the scale size of individual units to match better their local market conditions. The assumptions and structure of the aggregate and site-specific market efficiency models do not provide any "realistic" scale adjustment targets.

The improved scale-size (ISS) model was introduced which estimates targets altering the scale-size of controllable inputs/outputs of DMUs. The market efficiency targets obtained apply to management responsible for planning decisions within the MUO. The real life application of the ISS model demonstrated noticeable differences in its results as compared to those obtained from the aggregate and site-specific market efficiency models. The different results are particularly useful as they apply to different tiers of management within the MUO and indicate how different modelling assumptions can be used to encompass the multi-level structure of organisations in assessing performance.

The viability of MUOs is determined by the ability of individual units to generate profits. A systematic framework was developed seeking to investigate the association between DMUs' market efficiency and profitability. Four viability-quadrants can be developed with different policy implications for their corresponding units. The assessment of units' viability is, undoubtedly, very sensitive to criteria used to support the investigation and this resulted into defining variants of ***short and long run viability***.

The assessment of outlet viability was further explored using the Profits Impacts on Scale-Size (PRISS) regression based system. This system provides decision support concerning the profit implications of possible scale-size adjustments to controllable inputs/outputs of inefficient units. The application of PRISS to real data gives useful results that can be used to guide decision making at the DMU level.

Finally, the chapter sought to assess the *service-mix effectiveness* of individual DMUs. The methodology adopted revealed market efficiency differential due to different service mix which would have remained hidden otherwise.

- END OF CHAPTER EIGHT -

Chapter 9

Assessing marginal impacts of investments on the performance of organisational units[1]

1. Introduction

Chapter 9 addresses the issue of disentangling investment and environmental impacts on the comparative "market efficiency" of organisational decision making units (DMUs). In the previous two chapters a performance related decision support process was developed for profit making MUOs. The diagnostic part of this process (chapter seven) sought to assess market efficiency targets for different tier management. On the other hand, the planning part of the study (chapter eight) succeeded in estimating improved scale-size targets for scale inefficient units.

Central management of MUOs seeks to improve the performance of their network of DMUs by using capital investment projects for upgrading their outlets. This, it is hoped, should ultimately lead to higher corporate profitability. Central management needs to identify those DMUs which will secure the best return on funds invested. In the current practice, the identification of DMUs to receive investment funds often rests on little more than their profitability in the previous financial year. This, as was argued in chapter seven, is not a valid method of selection because it ignores the market efficiency of the DMU. A

[1] An earlier version of this chapter will appear in the International Journal of Production Economics (1995).

DMU showing a high profit could be benefiting from a very favourable environment and in fact securing only a fraction of its potential custom. In such a case further investment of funds would only be justified if they would lead to an improvement of the market efficiency of the DMU. In contrast a DMU enjoying a high market efficiency should only receive further investment funds if they are essential for the maintenance of its market efficiency.

Objectives of capital investment in MUOs relate to the specific characteristics of individual organisations and the nature of the sector in which they operate. Profit making MUOs invest to support their relative positioning within competitive markets seeking short/long run maximisation of corporate benefits. These objectives are operationalised using attributes such as sales' growth and increased market share; product diversification and improved quality of services; information technology and cost reduction investments.

The target setting process in either and diagnostic (chapter 7) or the planning (chapter 8) mode give estimates on what individual outlets should be achieving in principle had they operated efficiently. The management of the MUO is called upon to support the achievement of these targets by improving operating practices, changing the products it markets and increasing the marketing effectiveness of its outlets. In principle, most of these issues are not cost free and substantial capital investments are required for their implementation. For example, improving the site-specific market efficiency of outlets may require improved managerial practices by local management. The latter can be achieved by better training of new recruits and continuing training of current staff, by giving incentive plans and perhaps by employing management with higher qualifications. In the case of improved scale-size targets, again the achievement of the estimated targets is capital related as the expansion or contraction of the service area of outlets require substantial investment and/or relocation in an area with higher market potential.

Thus the likely impact of an investment of funds on the market efficiency of a DMU is of vital importance in the decision to invest or not funds in the DMU. It is not, however, straightforward to ascertain this impact. One possibility is to appraise changes in a DMU's market efficiency resulting from past investments made in it and extrapolate them to the future. However, this is not very simple. Most private sector organisations and many in the public sector operate in environments which constantly change. The number of drinks and meals sold over a period of time by a restaurant are affected by the general state of the economy and the same goes for the sales secured by other commercial outlets such as car showrooms and holiday outlets. Indeed, the same is true of public sector organisations selling goods and/or services. These changes in market conditions from one time period to the next compound the difficulty of prising out the effect of investments on the market efficiency of DMUs.

A variety of non-parametric models have been developed for assessing changes in efficiency over time. Chief among these are Charnes *et al.* (1985), Clarke (1992), Fare *et al.* (1989), Fare *et al.* (1992), and Tulkens *et al.* (1991). These models, however, cannot be used to assess the impact of investments on market efficiency because they do not allow for estimates of the impact of changes in market size over time to be incorporated within the model. These estimates are necessary if investment and market size effects are to be disentangled.

An approach is developed for assessing the impacts of investments on the market efficiency of DMUs. The investments considered relate to funds invested in the infra-structure of the DMU as distinct from investments in new products and/or services. Investments in infra-structure (e.g. in new furniture and fittings) impact indirectly on sales and the revenue they generate cannot be estimated by conventional accounting methods. The method rests on estimating the market efficiency of the DMU as it would have been without the investments and then contrasting that efficiency rating with its actual efficiency rating with the investments. Any difference in these two efficiency ratings reflects in part the effect of the investments on the market efficiency of the DMU.

Our attempt to disentangle market and investment impact on market efficiency has similarities in aim with Charnes *et al.* (1981) where an attempt was made to disentangle managerial and programme effects on efficiency. The approach adopted in our case is, however, different. In principle DMUs with investment can be classified in one programme and those without in a separate programme. However, this will still not make it possible to disentangle market and investment effects. For this estimates of the effects on market efficiency emanating from changes in market size alone are needed. Such estimates are used in the approach adopted here.

The structure of the chapter is as follows:

The next section develops the method for decomposing changes in market efficiency between two time periods into those attributable to investment and those attributable to changes in market size and other effects. The third section constructs the specific input-output set and associated DEA models for disentangling the impacts on market efficiency attributable to investments made in the pubs used already in the previous chapters. The final section discusses the results obtained.

2. Isolating the effects of investment on market efficiency

A key feature of market DMUs is that the volume of goods or services they sell is strongly influenced by the size of the market in which they operate. The size of markets generally changes over time. Consequently if we wish to assess the impact of an investment on the volume of goods or services sold by a DMU over time we need to segregate the impact of the change in the size of the market in which the DMU operates from the impact of investments made in its infrastructure. This section develops a method for this purpose.

2.1 Isolating investment impacts on market efficiency

Consider a set of j=1,...,n DMUs over a time period t=1,...T. For each time period t a 'production possibility set' Φ^t can be defined as follows:

$$\Phi^t = \left\{ \left(X^t, Y^t\right) \middle| \sum_{j=1}^{n} \lambda_j X_j^t \leq X^t \ ; \ \sum_{j=1}^{n} \lambda_j Y_j^t \geq Y^t \ ; \ \sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \geq 0 \right\}, \ t = 1, \ldots, T \quad \text{(M9.1)}.$$

Φ^t consists of all input levels X^t and corresponding output levels Y^t feasible in principle. The vectors (X_j, Y_j) consist of the vector $X_j \in \Re_+^m$ of input levels and the vector $Y_j \in \Re_+^s$ of output levels observed at DMU j and the superscript t identifies the time period concerned. This definition of Φ^t assumes variable returns to scale hold in the production process operated by the DMUs (see chapter 2). In the case of market DMUs inputs relate typically to market size and capacity of the DMUs while outputs relate to revenue generated (see in chapter seven and eight the set of intrinsic and extrinsic inputs used to describe the production technology of pubs).

The market efficiency of DMU j in period t is measured by E_j^t where

$$E_j^t = \max\left\{z^t \middle| X^t \leq X_j^t, Y^t \geq z^t Y_j^t, (X^t, Y^t) \in \Phi^t\right\} \qquad \text{(M9.2)}.$$

The assessment of market efficiency adopts an output expansion strategy (see chapters 7 and 8) as inputs in the study are the resources committed and outputs the sales generated by individual DMUs.

Let $\left(X_{cj}^t, X_{fj}^t, Y_j^t\right)$ be the input-output levels of DMU j at time period t so that X_{cj}^t are the levels of inputs which stay constant from t to t+1 when no investment in the DMU is made in year t, X_{fj}^t are the levels of inputs that generally alter from year t to t+1 irrespective of investments made in year t (for example the market size in which a DMU operates) and Y_j^t are the levels of outputs in year t. Finally let E_j^t be the market efficiency of DMU j computed using M9.2.

Let us assume now that an investment was made in DMU j_O at the end of time period t. We can now compute the following efficiencies:

(i). We can compute the market efficiency of DMU j_O, $E_{j_o}^{t+1}$ in period t+1. $E_{j_o}^{t+1}$ is measured using M9.2 with reference to Φ^{t+1} defined as in (2.1) using input-output levels observed in period t+1.

(ii). We can also compute the estimated market efficiency of DMU j_O, $\hat{E}_{j_o}^{t+1}$, in period t+1 had there been no investment in it at the end of period t. $\hat{E}_{j_o}^{t+1}$ is measured using M9.2 with reference to a new production possibility set $\tilde{\Phi}_o^{t+1}$. $\tilde{\Phi}_o^{t+1}$ is constructed as Φ^{t+1} in (i) above except that it contains the additional "DMU" which is being assessed and has the estimated input-output levels $\left(X_{cj_o}^t, X_{fj_o}^{t+1}, \phi^{t,t+1} Y_{j_o}^t\right)$ DMU j_O would have had in t+1 <u>without investment</u>. How these estimates might be obtained is discussed later.

The following measures can now be defined to reflect the effects on the market efficiency of DMU j attributable to investment and changes in the size of the market in which it operates:

The ***overall*** change in the market efficiency of DMU j between periods t and t+1 is reflected

$$O_j^{t,t+1} = E_j^{t+1} / E_j^t . \tag{M9.3}$$

The change in the market efficiency of DMU j between periods t and t+1 attributable to the change in the size of the ***market*** in which it operates *and to other factors* not related to investment is reflected in

$$M_j^{t,t+1} = \hat{E}_j^{t+1} / E_j^t \cdot \tag{M9.4}$$

The change in the market efficiency of DMU j between periods t and t+1 which is attributable to ***investments*** made in it in period t is reflected in

$$I_j^{t,t+1} = E_j^{t+1} / \hat{E}_j^{t+1} \cdot \tag{M9.5}$$

We shall refer to $I_j^{t,t+1}$ as the "<u>investment effectiveness</u>" of DMU j from period t to period t+1.

Note that $O_j^{t,t+1} = M_j^{t,t+1} x\, I_j^{t,t+1}$ so that the measure reflecting overall change in market efficiency is multiplicatively decomposed into a measure reflecting change in market efficiency mainly due to market size changes and the investment effectiveness measure. **The estimation and use of the investment effectiveness component $I_j^{t,t+1}$ is one of the main objectives of this chapter.**

2.2 Practical aspects of the method developed

Both E_j^{t+1} and $\hat{E}_j^{t+1}$ are defined on production possibility sets which depend on observed input-output levels of DMUs in period t+1. These levels in turn depend on investments made in DMUs <u>other</u> than j during time periods t, t-1 etc. Thus the measures

$O_j^{t,t+1}, M_j^{t,t+1}, I_j^{t,t+1}$ are relative to investments made in all DMUs in periods up to and including t.

This observation has important consequences for the use of the measure of investment effectiveness $I_j^{t,t+1}$ developed. The measure is seen as our best estimate of the effect on the market efficiency of DMU j due to the investments made in it, given the investments made in other DMUs todate. It is entirely possible that this estimate would be different had investments in DMUs other than DMU j been different to those actually made so far. It is, however, more sensible to use actual rather than hypothetical investments to estimate the impact on the market efficiency of DMU j from the investment made in it.

The key to measuring the investment effectiveness of a DMU is the estimate of the impact of market size change on its output levels. It is generally safe to assume that without investment a DMU would continue to use the same operating practices in the short term. So the major difference in its ability to generate outputs (notably revenue) between successive time periods when no investment is made will be due to changes in the size of the market in which it operates. There will, of course, be other random impacts on its outputs not related to market size changes but these are likely to prove minor in relation to those induced by changes in market size. Hence to estimate outputs as they would be without investment we need to focus on an estimate of the impact of market size changes on the DMU's outputs.

The simplest estimate of an output level at DMU j when no investment is made in the DMU is obtained by assuming that the output level would have altered between periods t and t+1 by the same factor as the size of the market alters between t and t+1 in the output concerned. For example in the application of the method to public houses outlined later, it was assumed that a public house ("pub") that had no investment between years t and t+1 would have experienced the same change in its sales of beer as was experienced in its catchment area overall. This is an acceptable basis for estimating the effects of market size changes when such changes are not affected significantly by any investments in the individual DMU concerned. (Investments in individual pubs, as pointed out later, generally affected only marginally the size of the market in their catchment area.)

The estimation of the investment effectiveness (defined in (2.5)) of a DMU with investments makes use of unobserved input-output levels. These are the input and output levels that would have been expected at the DMU without investment.

This has similarities with the so-called 'counterfactual' studies often found in economics. Counterfactual studies use econometric models to link input-output factors of economic systems where the levels of some of the factors have not actually been observed. The models are then used to estimate the effects of unobserved factor levels on the economic system. For example, using counterfactual studies Dimsdale and Horewood (1992) examine the effects of hypothetical public spending programmes on unemployment levels in the UK during the interwar period. In assuming the absence of investment when estimating input or output levels of a DMU with investment, the estimates are being derived in a counterfactual manner.

The use of unobserved output levels for estimating the investment effectiveness has also similarities with the concept of *latent* variables in econometric theory. Latent variables can be used to account for measurement errors in independent variables but more importantly for estimating the effects of unobservable factors on social and economic phenomena. For instance, Lanen *et al.* (1992) use latent variables to examine how adoption of contracts for executives affect performance of electric utilities in the USA. In estimating investment effectiveness in the case of pubs the *latent effect* is the *overall change in drink and food sales from time period t to t+1 in the market in which pubs operate.*

In the general multi input/output case the vector $\phi_j = \left(\phi_{x_j}, \phi_{y_j}\right)$ where $\phi_{x_j} \in \Re^m_+, \phi_{y_j} \in \Re^s_+$ can be used to estimate $\left(\hat{X}^{t+1}_j, \hat{Y}^{t+1}_j\right) = \left(\phi_{x_j} X^t_j, \phi_{y_j} Y^t_j\right)$, the input-output levels of DMU j in period t+1, without investment in period t. The vector ϕ_j can be estimated in several ways. Use can be made of general economic growth indices such as consumption and trade indices reflecting market size changes over time. Indices reflecting local market effects on specific inputs/outputs will also be needed. Alternatively, an attempt can be made to estimate the elements of ϕ_j by using causal regression or other more advanced stochastic methodologies along the lines of *latent* analysis (see Bollen (1989)).

The method developed in this paper makes it possible to compute the investment effectiveness of outlets for each year. This can be used to identify outlets and managements that offer the best scope of returns to investments.

The method is intended to be an aid rather than a sole instrument by which investments are decided. In particular, the method cannot direct investments to outlets that have had no investment in the past. However, for on-going organisations the proportion of outlets which never had an investment should be very small and therefore this should not present serious difficulties. The next section illustrates the use of the method in the brewery controlling the pubs used in chapters seven and eight.

3. Disentangling market and investment effects on market efficiency: An illustration using data on pubs

Capital investment plays a key role within the brewing industry in the UK. This is mainly due to the intensified competition and structural changes experienced currently within the industry in the UK. The general decline in beer consumption experienced in the UK market at the time of writing had profound effects on the investment strategies adopted by individual brewers. Some brewers concentrated their investment activities towards beer production whilst others emphasised investments on their network of pubs and restaurants. Investments on refitting and improving pubs and restaurants grew from £ 624 million in 1986 to £1131 million in 1990, Key Note (1991). For the individual brewing company was very important, therefore, to ensure that investment decisions are made to support the long run viability of the company. Assessment of the investment effectiveness of past investments in pubs would have particular importance in evaluating how "effective" decisions made in the recent past were.

As noted earlier the brewery owns several hundred pubs (2,000 pubs) but the study focused on a subset of 154 pubs for which all the data required existed. Some 78 of these pubs had had capital invested in the period 1988-1991 and this was the period covered by the study. No pub had more than one investment of capital in the period under consideration.

Corporate management identifies the outlets to receive investment funds and the nature of the investment to be made in each case. Managers of individual pubs have no discretion on how investment funds are used.

Table 9.1 summarises the amounts invested in the four year period 1988-1991 in pubs within the set of 154 pubs being studied. The Table also shows the numbers of pubs that received investment funds each year. Amounts shown are in real (1988) prices.

Table 9.1

Capital Investment in Public Houses
(£ '000 per pub that received investment that year)

YEAR	Pubs with investment	Average per pub	Maximum investment	Minimum investment
1988	7	137763	191748	61652
1989	18	146384	253342	63868
1990	26	147252	339303	49712
1991	27	147675	554534	34644

The amount invested each year has on average remained relatively stable after a rise from 1988 to 1989. Total funds invested have, however, risen steadily as more and more pubs received investment funds. It is noteworthy that maximum amounts invested rose too over time representing more substantial investments in the cases of certain pubs.

Estimates of $I_j^{t,t+1}$ and $M_j^{t,t+1}$ (see M9.5 and M9.4) were obtainable only for investments made during the period 1988-1990 as their estimation requires observations of at least one year after the investments have taken place. There were 51 pubs that had investments in these three years but data for all four years was available for only 42 of them. The effects of investment on the market efficiency of these 42 pubs were studied.

The input-output variables selected for assessing the market efficiency of the pubs were the ones selected in the diagnostic and planning analysis of chapter seven and eight. This input/output set in listed in Table 9.2 for completeness.

Table 9.2

Variables for Assessing the Market Efficiency

INPUTS	OUTPUT
Bar area (ft^2)	Turnover from the sale of food and drinks (£)
State of repair	
No. of car park spaces	
No. of competitors	
Consumption of alcohol in catchment area (barrels of beer)	
No. of potential customers in catchment area	
Households' income in the catchment area	

The catchment area of a pub is taken as the area within a 2.5 miles radius from the pub.

The variables in Table 9.2 were arrived at after much deliberation and discussions with corporate management. Pricing, target clientele and investments at a pub are the preserve of the controlling brewery. The assessment therefore seeks to reflect the efficiency of pub management in attracting custom given the decisions made centrally by the brewery about the pub's operations.

The input variables can be classified into *internal* and *external.* Internal variables reflect factors decided upon by the brewery (i.e. location, state of repair, bar area, and car park facilities). External variables reflect environmental factors outside the brewery's control. None of the input variables is controllable by pub management.

3.2. The DEA models used

Using M9.2 the market efficiency in year t+1, $E_{j_0}^{t+1}$, of pub j_0 which had an investment in year t is

Market efficiency model (M9.6)
$E_{j_o}^{t+1} = \max_{\lambda_j, z} \quad z$
$s.t. \sum_{j=1}^{n} \lambda_j x_{ij}^{t+1} \le x_{ij_o}^{t+1} \quad \forall i$
$\sum_{j=1}^{n} \lambda_j y_j^{t+1} \ge z y_{j_o}^{t+1}$
$\sum_{j=1}^{n} \lambda_j = 1$
z free and $\lambda_j \ge 0$, $\forall j = 1, \ldots, n$

The formulation in M1 is based on the DEA models developed by Banker and Morey (1986b) for treating uncontrollable inputs. (The Banker and Morey (1986b) model (19) pp. 516 reduces to (M1) when, as in our case, all inputs are uncontrollable and an output orientation is used.) x_{ij}^{t+1} and y_j^{t+1} represent the ith input and output level of pub j in year t+1.

The market efficiency model in M9.6 is what was called as "*site-specific market efficiency*" of outlets in the diagnostic study of chapter seven. The decision to use the local market efficiency of pubs for assessing the *investment effectiveness* avoids penalising pubs which operate at an unproductive scale size. Only by removing the scale bias from the assessment of market efficiency shall we be able to answer the question of what the market efficiency of pubs would have been had no investment taken place.

The estimated market efficiency $\hat{E}_{\hat{j}_o}^{t+1}$ of pub j_O during year t+1, had it had no investment at the end of year t, was assessed by modifying M9.6 above to model M9.7. 'Pub j_O' without investment is labelled as $\hat{j}_o$.

Market efficiency free of investment effects (M9.7)

$$\hat{E}_{\hat{j}_o}^{t+1} = \max_{\mu_j, q} \quad q$$

$$s.t. \quad \sum_{j=1}^{n} \mu_j x_{i_c j}^{t+1} + \mu_{\hat{j}_o} x_{i_c j_o}^{t} \quad \leq \quad x_{i_c j_o}^{t} \qquad \forall i \in s_c$$

$$\sum_{j=1}^{n} \mu_j x_{i_c j}^{t+1} + \mu_{\hat{j}_o} x_{i_f j_o}^{t+1} \quad \leq \quad x_{i_f j_o}^{t+1} \qquad \forall i \in s_f$$

$$\sum_{j=1}^{n} \mu_j y_j^{t+1} + \mu_{\hat{j}_o} \left(\phi_{j_o}^{t,t+1} y_{j_o}^{t}\right) \quad \geq q \left(\phi_{j_o}^{t,t+1} y_{j_o}^{t}\right)$$

$$\sum_{j=1}^{n} \mu_j \quad + \mu_{\hat{j}_o} \quad = \quad 1$$

$$q \text{ free and } \mu_j \geq 0 \ \forall \, j = 1, .., j_o, .., n, \hat{j}_o$$

Where x_{ij}^t and y_j^t are as defined for M9.6, $\phi_{j_o}^{t,t+1} y_{j_o}^t$ is the estimated turnover of pub j_o during year t+1 had it had no investment during year t. The next subsection outlines the method followed for computing these estimates. s_c and s_f are the subsets of inputs whose levels change and do not change respectively due to the investment in period t. These subsets generally vary with investments.

The inclusion of pub $\hat{j}_o$ in the set of "observed" input-output levels may at first appear odd. However, if we assume that our estimated input-output levels of pub j_o during period t+1 reflect what would have held true had pub j_o had no investment during period t then there is no reason to exclude these estimates from the input-output levels used to define the production possibility set for period t+1. (The interested reader is referred to further discussions at this point in Berg *et al.* (1992) and Elyasiani *et al.* (1990).)

3.3. Estimating the input-output levels of a pub

As noted above, it was necessary to estimate for the purposes of model M9.7 the levels of the inputs and the output of pub j_o during period t +1 as they would have been had there been no investment in the pub during period t.

The internal variables were not affected by the investments during the period considered except for the state of repair. This was only available for 1991, that is as it was <u>after</u> the investments had been made. This changed the interpretation of the efficiency $\hat{E}_j^{t,t+1}$ obtained from M9.7. $\hat{E}_j^{t,t+1}$ is now an estimate of the market efficiency of pub j in year t+1 as it would have been with investment in year t had its turnover in t+1 merely reflected overall market size changes. Thus, $I_j^{t,t+1}$ as defined in M9.5 still reflects the extent to which investment has enabled pub j to perform better than if it was merely following market size changes.

The output level (turnover) $\phi_j^{t,t+1} y_j^t$ for year t+1 net of the effects of investment in year t was estimated as follows.

Let y_j^t be the turnover of pub j during year t, y_{Dj}^t its revenue from the sale of drinks and y_{Mj}^t its revenue from the sale of meals during the same year. Then we have $y_j^t = y_{Dj}^t + y_{Mj}^t$. The sales of drinks $\hat{y}_{Dj}^{t+1}$ and meals $\hat{y}_{Mj}^{t+1}$ as they would have been without investment in pub j in year t were estimated as follows:

$$\hat{y}_{Dj}^{t+1} = \frac{TD_j^{t+1}}{TD_j^t} y_{Dj}^t \qquad \text{(M9.8)}$$

and

$$\hat{y}_{Mj}^{t+1} = \frac{TM_j^{t+1}}{TM_j^t} y_{Mj}^t \qquad \text{(M9.9)}$$

where

TD_j^t and TM_j^t represent the total sales of drinks (barrels) and number of meals served in the surrounding area of pub j in year t.

y_{Dj}^t and y_{Mj}^t were adjusted for inflation using the drink and food retail prices indices so that all monetary values used in M9.8 and M9.9 were constant at 1988 values. This ensures year on year comparisons reflect revenue changes net of inflation.

The estimated output level $\phi_j^{t,t+1} y_j^t$ for pub j during period t+1 without investment in year t was

$$\phi^{t,t+1} y_j^t = \hat{y}_{Dj}^{t+1} + \hat{y}_{Mj}^{t+1} \qquad \text{(M9.10)}$$

The estimated turnover in pub j using M9.10 assumes that without investment in year t the pub would have registered in its drinks and meals sales the same change as did the size of its market in these two areas.

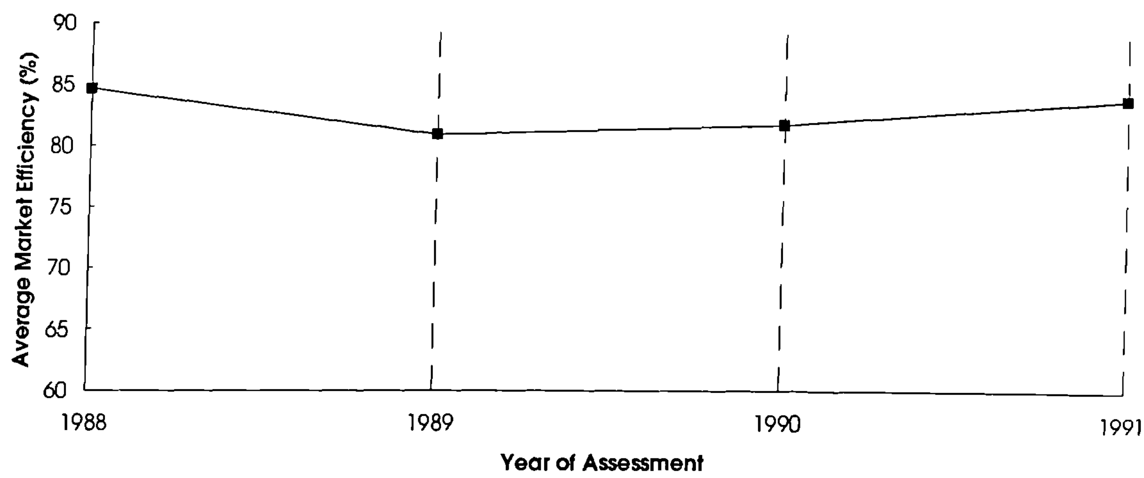
All investments were completed within the financial year in which the funds were allocated so that their effects on market efficiency should be apparent in a subsequent year. Disruption in the year of investment t could, however, lead to $\phi_j^{t,t+1} y_j^t$ being an underestimate of potential sales in year t+1. This in turn can lead to an overestimate of $I_j^{t,t+1}$.

4. Results

Before looking at the investment effectiveness of pubs, their average market efficiency during each one of the four years 1988-1991 was computed using model (M1). The aim was to obtain an overview of how average market efficiency has changed in the light of continuing investments in the pubs. The results appear in Figure 9.1, where efficiencies defined in M9.6 have been inverted to give $100/E_j^t$. This means the efficiency ratings in Figure 1 represent the percentage of its potential revenue a pub realises.

Figure 9.1

Average site-specific market efficiency of pubs



Average efficiency is relatively stable during the four-year period, if somewhat down. This means that either investments have not narrowed in any way the average gap between the proportions of their potential revenue the pubs actually realise or investments have revealed an increasing potential for revenue generation. If potential revenue has risen then stable mean market efficiency means rising mean gross revenue. In the case of the pubs, however, mean gross revenue declined through recession. This still, however, cannot be taken to mean that investments have been ineffective. The revenue realised might have been lower still without investments.

The investment effectiveness indices $I_j^{t,t+1}$ of the pubs are summarised in Table 9.3.

Table 9.3

Investment Effectiveness Indices

Years of assessment (t to t+1)	Number of pubs with $I_j^{t,t+1} \geq 1$	Number of pubs with $I_j^{t,t+1} < 1$	$I_j^{t,t+1}$ Qr1	$I_j^{t,t+1}$ *median*	$I_j^{t,t+1}$ Qr3
1988 to 1989	4	2	1.00	*1.03*	1.04
1989 to 1990	11	8	0.93	*1.02*	1.07
1990 to 1991	28	14	0.94	*1.05*	1.08
Total	43	24	0.94	*1.03*	1.07

Qr1 and Qr3 are respectively the first and third quartile. Note that pub numbers are cumulative.

The following observations can be made:

- *The market efficiency of pubs that have investments is generally 2 to 5 percentage points higher than would be expected if their turnover had merely reflected overall market changes.*

This can be deduced from the median value of $I_j^{t,t+1}$ in Table 9.3. The rise in market efficiency appears managerially if not statistically significant. The improved market efficiency of pubs with investments can naturally not be taken at face value. The estimated turnover at a pub is subject to uncertainty. Also there could be factors unrelated to investments that have generally influenced positively the market efficiency of pubs with investments. The approach does, however, at least indicate where investment of capital might have led to better sales than would otherwise have been the case.

- *Consistency of investment effectiveness is variable.*

Eighteen of the 42 pubs showed consistently investment effectiveness above 1. Five had consistently investment effectiveness below 1. The remaining pubs had inconsistent results, with investment effectiveness some of the years above and the rest below 1.

- *Investment effectiveness was associated with location of pubs*

Pub locations were characterised suburban, main road or industrial estate. The null hypothesis of no association between pub location and investment effectiveness was rejected at the 5% level of significance. Pubs located in suburbs appeared more likely to register better than expected market efficiency. Thus, more generally the analysis can lead not only to the identification of individual DMUs but also to environments where investments might be more effective in securing better than expected market efficiency.

- *There are significant differences in the investment effectiveness of individual pubs.*

The interquartile range of the investment effectiveness in Table 9.3 is quite large, covering generally some 14 percentage points. Pub 19, registered the largest investment effectiveness. It was from the year 1990 to 1991. Investment in it in 1990 was above average but substantially below the maximum size of investment in that year. At the other extreme Pub 5, registered the lowest investment effectiveness at 0.66, from 1989 to 1990. It had an investment of average size. Clearly in both cases the size of the investment does not seem likely can explain the investment effectiveness observed. There may, of course, be factors not reflected in the input-output variables used which can explain the apparent good performance of Pub 19 and weak performance of Pub 5. The usefulness of the analysis lies in identifying pubs such as Pub 5 and 19 which exhibit performance substantially at variance with expectations. Their operations can be further analysed outside the DEA context to identify the factors responsible and use the information to improve performance of pubs generally.

- *There are indications that market effectiveness is highest immediately after an investment is made.*

This can be seen more clearly in Table 9.4 where the investment effectiveness ratio computed one, two and three years after the investment is shown. Note that $I_j^{t,t+1}$ is always computed for consecutive years but the investment itself could have occurred before t.

Table 9.4
Investment effectiveness $I_j^{t,t+1}$

Years between the year of investment and t+1	Qr1	Median	Qr3	Sample size
1	0.94	1.046	1.08	42
2	0.86	1.02	1.065	19
3	0.95	0.96	1.03	6

The sample where there has been a third year of operation after investment is too small to be reliable. However, the sample sizes for 1 and 2 full years of operation after the investment are fair and it does appear that market efficiency is higher than expected in the year following immediately the one in which an investment was made. This could be simply because disruption in the year of investment has led to an underestimate of market efficiency as it would be the following year without investment. It could, however, also be indicating that the life of investment in general infrastructure, (refurbishment etc.) is relatively short and continuing investments are necessary if pubs are to progressively improve on market efficiency. Better estimates of sales due to market size changes alone could help resolve this uncertainty.

- *Pubs that show high profitability and market efficiency prior to investment are likely to show high investment effectiveness.*

This can be deduced by looking at the results in Table 9.5 It shows the median market efficiency and profitability of pubs of each category before and after investment and of their investment effectiveness.

Table 9.5
Investment effects for pubs of each category

Groups of Pubs		Prior to Investments (median)		After the investments (median)	
Category	No. of Pubs	Market Efficiency	Profitability	Investment effectiveness ratio	Profitability
"?"	6	60.5	0.18	0.95	0.24
"Sleepers"	9	56.2	0.30	0.99	0.25
"Stars"	11	100	0.34	1.06	0.335
"Dogs"	16	100	0.24	1.005	0.19

The "*stars*" are pubs with high market efficiency and profitability. Pubs in the "?" category have low profitability and market efficiency. "*Sleepers*" have high profitability but low market efficiency. Finally, "dogs" have high market efficiency but low profitability.

Pubs in the "?" category (low efficiency and profitability) appear to show low investment effectiveness. However, their median profitability rises substantially after investment from 18% to 24%. Although the sample is small one likely explanation is that these pubs faced with decreasing revenues showed better cost control which resulted in higher profitability. The indications are, however, that despite investment they are not exploiting their market potential in full.

Pubs in the sleeper and dogs categories appear to offer no better market efficiency than might be expected without investment. Their investments effectiveness is about 1 but their profitability declines. Possibly no action was taken to contain costs in a decreasing gross revenue situation during the period covered by the study.

Stars (high market efficiency and profitability prior to investment) appear to offer the highest investment effectiveness. Their profitability is virtually unchanged before and after investment but the improved market efficiency should result in higher gross profits (if the market size is increasing) than would be the case without investment. This tends to justify selecting star pubs for further investment. Clearly there is a case of studying operating practices at "star" pubs and transferring them to the rest of the pubs.

5. Conclusions

Investment in the infrastructure of trading outlets is an important part of their short term management. In this chapter a method was developed to aid corporate management to select recipients of infrastructure investments.

It was argued that it is not sufficient to merely contrast the market efficiency of a DMU before and after investment in order to determine how effective investment in its infrastructure has been. An allowance needs to be made for the fact that market DMUs operate in changing market sizes and the effects of these changes need to be disentangled from those of investments. The method developed relies on contrasting the market efficiency at a DMU after investment with its market efficiency as it would have been without investment. In this way an estimate was made of the improvement of the market efficiency of a DMU which may be attributed to the funds invested in it.

An organisation controlling a set of market outlets would normally have data on investments made on outlets in the past. The method developed in this paper makes it possible to compute the investment effectiveness of such outlets for each year. The information can be used in a number of ways:

- To identify outlets and managements that offer the best scope of returns to investment
- To estimate how rapidly investment effectiveness wears off after an investment is made.
- Environmental features conducive to good investment effectiveness.

The method is intended to be an aid rather than a sole instrument by which investments are directed. In particular, the method cannot direct investments to outlets that have had no investment in the past. However, for on-going organisations the proportion of outlets which never had an investment should be very small and therefore should not present serious difficulties for the use of the method developed.

The method developed was illustrated using data from a set of 154 pubs. Data was available on investments made in these pubs in the four year period 1988-1991. It was found that generally investment increased market efficiency. This, however, appears to depend on location of the pub, as well as on their market efficiency and profitability prior to investment. Pubs with high efficiency and profitability before investment were more likely to improve their market efficiency while pubs with low efficiency and profitability appear to suffer lower market efficiency after investment.

The estimates of the effects on market efficiency that the method provides are subject to uncertainty. However, the method is useful at least for indicating those DMUs that are likely to secure a good return on market efficiency if further funds are invested in their infrastructure.

- END OF CHAPTER NINE -

Chapter 10

Conclusions and further research avenues

1. Introduction

This chapter reviews the ideas and models put forward in the previous nine chapters of the thesis and discusses some possible directions of future research. Decision support is undoubtedly a critical factor of effective decision making. The complexity multi-unit organisations (MUOs) face in their operations requires the development of sophisticated systems that would enable management to *control* current operations and *plan* for the future. As it was noted in chapter one, managerial control and planning are usually separate processes in organisations. Yet there is widespread appreciation of the potential benefits from opening communication links between the two processes. This thesis was motivated by the lack of adequate operational models to address the question of linking control and planning mechanisms and thus new approaches have been developed which:

> **seek to co-ordinate management control and planning at individual activity centres of MUOs in order to support their operations towards responding successfully to their market competition (profit oriented) or to the demand for services (not-for-profit).**

The thesis has distinguished two broad types of resource allocation problems incurred in MUOs.

The **first** concerns activity centres that manage their budgets without interference from their headquarters. This has been characterised as *a-posteriori* decision making since the use of resources is decided after the resources have been allocated to individual units. The **second** problem concerns the allocation of resources to activity centres to realise specific projects decided upon by headquarters. This has been characterised as *a-priori* decision making since the allocation of resources is project specific and projects are centrally selected.

The decision support framework for resource allocation consists of *diagnostic* and *planning* phases. The diagnostic phase focuses on the extent to which DMUs support adequately the objectives of the central organisation. The planning phase on the other hand draws on the findings of the diagnostic process and develops scenarios of alternative courses of action for improving the performance of the organisation.

The thesis has proposed a series of analytical models linking control and planning mechanisms. In this concluding chapter the intention is to summarise the rationale of these models. The chapter starts by discussing the rationale of the centralised and decentralised models for target setting and resource allocation in the *a-posteriori* decision making case. A similar discussion follows on the diagnostic and planning support system for the *a-priori* decision making case. The chapter concludes by giving directions for future research concerning the development and use of frontier analysis as a decision support tool.

2. Decision support for *a-posteriori* diagnosis and planning

A-posteriori diagnosis and planning takes place at MUOs which allow their individual units to deploy their resources without direct interference from central management. However, as organisations operate with scarce resources and intense pressure for accountability, it is necessary to introduce mechanisms for monitoring resource utilisation. The need to develop and use effective control mechanisms is greater because of the limited involvement of central management in the actual use of the resources. The obvious paradox is that the higher the discretion of activity centres to make decisions the higher is the need for central management to assess performance effectively.

The development of a-posteriori decision making mechanisms in chapters three to six was pursued having as objectives:

- To disentangle the concept of target setting from that of efficiency assessment.
- To link *control* and *planning* mechanisms, using as a vehicle the target setting process.

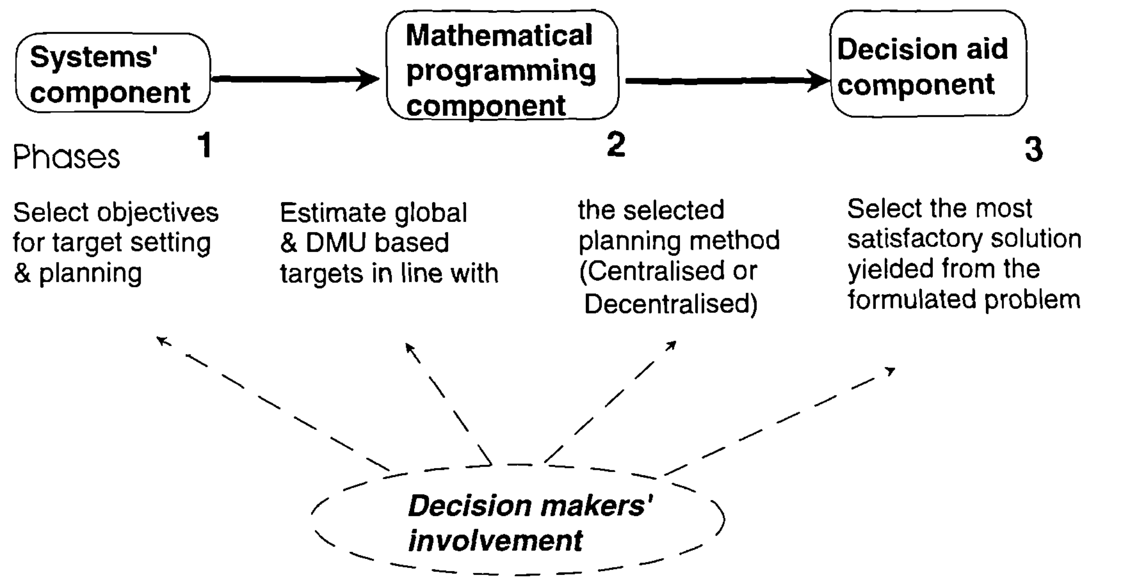
- To develop centralised and decentralised decision support systems which assess targets and allocate resources as parts of a simultaneous process.

The notion of target setting in MUOs was strengthened in order to be used as a control and planning mechanism. Principles of effective target setting were posited and used to develop the target setting model M4.3 in chapter four. This model was an extension of earlier work done by Thanassoulis and Dyson (1992) for estimating performance targets. It enables the target setting process to consider as alternative courses of action the possibility of increasing the resources used by decision making units (DMUs) anticipating higher output returns or reducing the outputs delivered anticipating more than proportionate input savings. Chapter four also develops a model for setting performance targets at the global organisational level in the form of model M4.4. The target setting framework developed, in Chapter four, has a number of unique features including :

- The active involvement of decision makers during the target setting process due to its interactive-iterative nature,
- The use of input/output substitutability in the process of setting targets.

Despite the advanced features of the target setting models M4.3 and M4.4 the question of resource allocation and decision support cannot be addressed in full since the target setting models focus on the performance of one unit at a time without considering interactions among them. These issues were the main focus of chapters five and six where the *a-posteriori* decision support framework was developed. A pictorial representation of this framework is given in Figure 10.1.

Figure 10.1
Decision support for target based planning



The Decision Support System (DSS) in Figure 10.1 is made of three interrelated phases which are discussed in more detail next.

2.1. DSS: phase one

The systems component within the DSS encapsulates the modelling aspects of the process. That is, the conversion of the general organisational objectives into input-output models which are ultimately the basic components of the mathematical formulations that follow in the next phase of the DSS.

The role of the systems components has already been acknowledged in Chapter four as a necessary component in the development of performance measurement systems. This role is enlarged, however, in the development of the DSS. Decision makers and analysts are expected to reach agreements, concerning the nature of organisational objectives that need to be pursued via the planning models developed. This includes the identification of appropriate criteria to represent the three fundamental objectives, namely efficiency, effectiveness and equity, of resource management in MUOs.

2.2. DSS: Phase two

The second phase of the DSS process is the "engine" of the system as it includes the selection of mathematical models for estimating performance targets at the DMU and global organisational level in line with the centralised or decentralised process of the organisation.

The mathematical models used for assessing performance targets seek to encapsulate the principles of effective target setting as developed in chapter three of the thesis. It is worth mentioning that the targets estimated reflect the decision makers preferences incorporated in the estimation process as illustrated numerically in the second half of chapter four.

The operationalisation of the linkage between target setting and resource allocation is the next issue of consideration. The problem of performance and resource allocation in multi-level multi-unit organisations (MULO) has been subject to substantial research effort by the management scientists and economists but as noted in chapter five limited progress has been made in linking resource allocation and target setting in real life problems. The goal programming formulations in chapters five and six were put forward for addressing these kind of problems. Two alternative formulations were used to accommodate the different behavioural issues arising in centralised and decentralised multi-level planning.

The distinction between *centralised* and *decentralised* target-based planning models goes beyond the differences concerning the solution process of multi-level planning problems. The main research stream, in multi-level programming problems, is based on the assumption that centralised and decentralised planning are alternative solutions processes for similar planning formulations (e.g. the *price* and *resource* directive methods discussed in chapter five fall on this stream). The centralised and decentralised models in chapters five and six adopt a different philosophy towards planning. The main differences lie on the way the organisational objectives are encompassed within the planning process and also on the extent to which the planning process is determined by the preferences of the management of the organisation. The discussion of the characteristics and differences between these models that follows is facilitated by the information provided in Table 10.1.

Table 10.1
Features of centralised and decentralised target-based DSS

Centralised model (M5.2)	Decentralised model (M6.4)
i. The **formulation** of the target setting models is made by central management.	i. Different management tiers are involved in the **formulation** & **estimation** of targets.
ii. **Implicit** representation of resource management objectives (3E's).	ii. **Explicit** representation of resource management objectives (3E's).
iii. Activity centres are expected to operate with efficient **technologies** selected centrally. (i.e. they are judged against a central system of values of inputs/outputs)	iii. The efficient technology of activity centres reflects their **own performance priorities** concerning the values of inputs/outputs.
iv. The solution process allows central management to select the most appropriate strategy for meeting the **global organisational targets.** Tradeoffs are considered among the decision makers of central management.	vi. Interactions with central management are required to resolve the **internal** and **external tradeoffs** that emerge from the maximisation of the resource planning objectives.
v. The solution process is **faster** at the risk, however, of **lower effectiveness** due to the lack of participation by the lower level of management.	v. The method is **computationally expensive** but the likelihood of reaching **viable** solutions is much higher due to the high degree of consensus that is reached.

The characteristics of centralised and decentralised DSS models in chapters five and six represent two alternative routes for solving target-based resource allocation problems. It is important to emphasise, however, that the two models are not the alternative solutions to the same problem as:

I. Centralised planning models seek to select the most favourable efficient technology for the inputs-outputs of individual DMUs in order to achieve the global organisational targets. The efficient technology of each DMU is selected globally and it reflects central management values on inputs/outputs.

II. Decentralised planning models, on the other hand, seek to maximise the global organisational objectives given the efficient technology of individual DMUs. The efficient technology of each DMU is selected separately and it reflects local management values on inputs/outputs.

Centralised and decentralised planning draws upon the fundamentals of the principal-agent theories. Van Ackere (1993), advocates the principal-agent relevance in production planning problems as a mechanism that can be used to facilitate optimal contracting relationships among the members of a firm where each member is motivated by self-interest. The models developed in this thesis go beyond the basic principal-agent frameworks as they seek:

- to incorporate strategic issues in planning,
- to setup contracts based on the selection of efficient technologies,
- to develop systematic ways for estimating performance targets for agents and principals,
- use of iterative/interactive processes which can be thought of as a means for renegotiating the contracts between principals and agents.

Van Ackere (1993), calls for an expansion of the current domain of principal-agent theory as a prerequisite for its effective application to management science problems. The planning models developed in the thesis represent a move towards that direction.

2.3. DSS: phase three

Figure 10.1 depicts the solution process of the DSS models formulated as an independent component. This was done intentionally in order to highlight the effects of the solution process on the planning scenarios developed. The target-based planning models in chapters five and six have multiple objectives. These objectives, namely efficiency, effectiveness and equity, are both in centralised and decentralised planning the same. However, the way they are addressed is different in each case.

Multiple objective programming problems have typically many satisfactory solutions and, therefore, one needs to consult the stakeholders of the modelling process (decision makers) for selecting the most preferred one. Typically, multiple iterations between analysts and decisions makers would be required before a final solution can be selected.

The solution of the centralised target setting model in M5.2 relies largely on the preferences and priorities of central management. This makes the solution process more efficient as the tradeoffs are constrained only within one level of the managerial hierarchy. The risk associated with this strategy concerns the lack of participation from the lower levels of management which is not allowed for by the model. In contrast, the decentralised planning model in M6.4 permits preferences and priorities both from central and local management in its solution process.

3. Decision support for *a-priori* diagnosis & planning

The second part of the thesis sought to aid resource allocation in *a-priori* decision making problems. The overall aim was, first to provide sufficient *diagnostic* information of the performance of individual DMUs, and second to guide management to select units to receive capital allocations. The models developed in the *a-priori* decision making case were tailored to multi-unit organisations operating in a market setting.

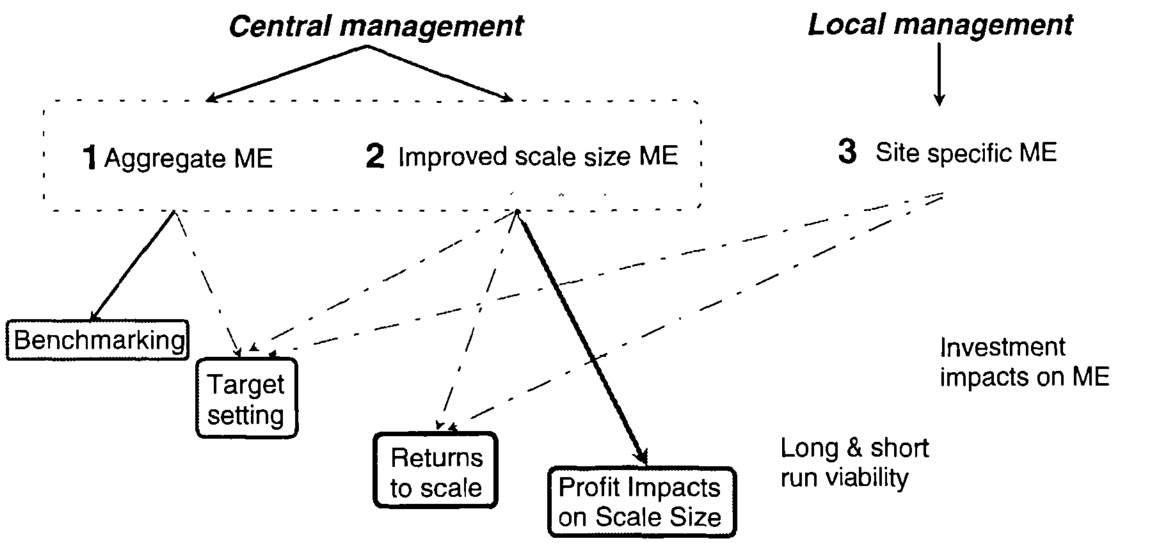
Profit making organisations deploy resources within their network of DMUs in order to maximise the penetration of their markets. Individual DMUs, despite the similar missions they pursue, have substantial differences in their size, scope and operating profile that create difficulties in assessing their performance. A two stage process was defined for assessing the *market* and *cost* efficiency of individual DMUs. Market efficiency assesses the relative success of individual outlets in attracting custom whilst cost efficiency concentrates on the outlets' ability to control costs.

Three definitions of market efficiency were developed reflecting the responsibilities of different tiers of management in MUOs.

- The ***Aggregate*** *market efficiency*, focuses on the control exercised by the central management of the organisation (e.g. location, input/output scale). Its assessment (model M7.1 in chapter seven) was pursued adopting a constant returns to scale economic assumption compounding elements of scale and management in the efficiency ratings obtained.
- The ***Site specific*** *market efficiency*, focuses on the performance of the management of individual DMUs (model M7.2 in chapter seven). The economic assumption of variable returns to scale was adopted here removing the scale bias from the assessed efficiency.
- The ***Improved scale-size*** *market efficiency* (ISS), seeks to estimate targets for scale size improvements as a vehicle for augmenting DMUs performance (model M8.3 in chapter eight). This type of market efficiency seeks to adjust the scale of controllable inputs/outputs of individual outlets and, therefore, an assumption of constant returns to scale is inherent in the assessed targets.

The three definitions of market efficiency (ME) were used as the basis for building a *diagnostic-planning* framework of decision support exhibited in Figure 10.2.

Figure 10.2
Diagnosis & planning models as decision aids for MUOs in market settings



Results obtained from the three market efficiency models were used in the diagnostic and planning analysis as follows.

Market efficiency targets were customised for different tiers of management in accordance with their control over the allocation of resources. These targets assessed the extent to which expected performance improvements of individual DMUs can be associated to different levels of management control. The market efficiency targets were estimated using relatively efficient DMUs which are considered as exemplary performers. Some additional tests were employed to identify those efficient DMUs which can be used by management as the *company benchmarks.*

The question of scale was given sufficient emphasis in both the diagnostic and planning components of the analysis. Results from the site-specific and aggregate market efficiency model were combined to investigate the extent to which individual outlets operate at a wrong scale. Further insights were obtained by the improved scale size (ISS) model that was used to derive scale adjustments for controllable inputs/outputs. The analysis also showed how sensitive the conclusions on the nature of economies of scale can be. For instance, when the site-specific market efficiency model was used to characterise economies of scale, a large proportion of a set of units (see Table 8.7 in chapter eight) were found to be operating under local increasing returns to scale. The opposite results, notably decreasing returns to scale, were found when the same units were assessed using the ISS model. This variation has profound managerial implications as, for example, using the site-specific model decision makers were advised to expand the scale size of DMUs whilst the ISS model advised reduction in the scale of controllable inputs. The conflicting results indicate how important the selection of efficiency models is particularly when is related with the provision of decision support.

It was suggested that the *long and short run viability* of individual outlets can be assessed combining information on external (market efficiency) and internal (profitability) performance indicators of the network of DMUs. The efficiency profitability matrix is used in chapter seven to classify DMUs in clusters with similar viability characteristics.

Further investigation of outlets' viability was made using the profit impacts of scale size (PRISS) regression based system in chapter eight. PRISS sought to provide an evaluation mechanism to assess the impacts of altering the scale of individual outlets on their profits. Application of the PRISS method on real life data showed, as expected, that DMUs that expand their scale size of operation (increasing returns to scale) would expect considerable profit increases.

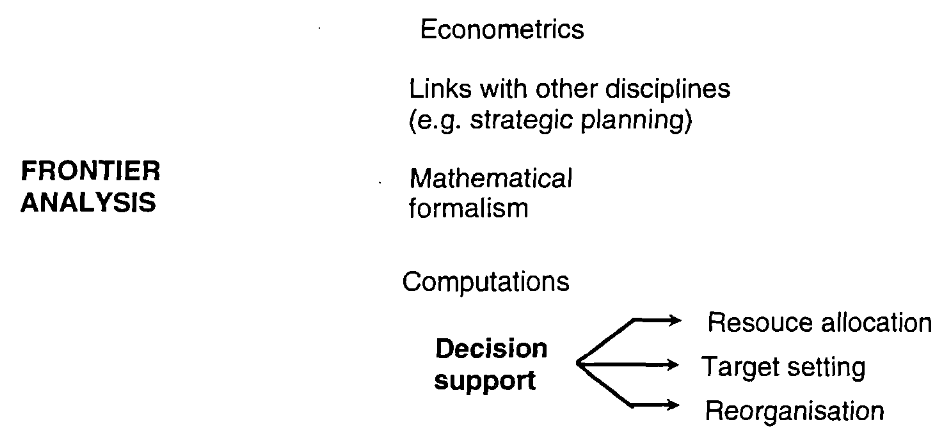
The *decision aid* component is the third part of the decision support process. *A-priori* decision making concerns the allocation of resources in MUOs, and relates to problems of capital investment and/or reorganisation of the services provided by DMUs. The contribution of the decision support mechanism developed in chapters seven to nine is primarily concerned with the identification of best operating practices, the estimation of targets for adjusting the scale size of operation of DMUs and the assessment of the impacts of past investments on the performance of units. This information can be used as the basis for improving the effectiveness of decision making tools used for allocating capital within the MUOs (e.g. project appraisal).

The provision of decision support for selecting projects to improve the operating profile of individual DMUs was pursued further in chapter nine. An evaluation process was put forward seeking to estimate the marginal impacts of past investments on the performance of DMUs. The *investment effectiveness* indices obtained from this new methodology can be used to guide management to invest in DMUs with the highest market efficiency returns.

4. Frontier analysis and decision support: avenues for further research

Frontier analysis as introduced in chapter two has been proposed as a tool for assessing the performance of activity centres drawing upon a wide spectrum of the political science, economics, management science and accounting literature. Since the original development of frontier analysis by Farrell (1957) and its operationalisation by Charnes *et al.* (1978) a considerable expansion[1] of the method has taken place. Nowadays one can safely talk about different research directions within the frontier analysis area. Figure 10.3 below exhibits a framework of the current development of independent research disciplines within the area of frontier analysis.

Figure 10.3
Frontier analysis & research directions



It is beyond the remit of this chapter to discuss all branches listed in Figure10.3 in full as the thesis is concerned with the decision support dimension of frontier analysis. These decision support mechanisms were developed in two different types of managerial problems (*need-based* and *market-based* performance measurement and resource allocation problems).

1 In one of the panel discussions focusing on modelling issues of DEA, organised in the EURO XII-OR36 conference in Glasgow (1994) Arie Lewin (Professor of Business Administration at the Duke University) paralleled the current development of DEA as very similar to the development and progress of regression analysis in the post second war period.

The research questions addressed in the thesis give rise to a number of issues that constitute themselves directions for further research. These research directions are related to the models and techniques developed in the thesis but also with more general issues regarding the future status of performance measurement as a part of the managerial process in multi-unit organisations.

4.1. Combining the target-based methods of the thesis with traditional resource allocation approaches

The establishment of efficiency as one of the key components of public management within the market economies is universally acknowledged. The assessment of efficiency, however, has so far been more prevalent in an auditing rather than planning context. There are recent moves, however, towards incorporating efficiency within the legislative processes of resource allocation. In the USA, for example, there is the issue of progressive reimbursement of hospitals on the basis of the costs of efficient benchmarks. An analogue, can also be found in the way resources and activities are allocated to individual hospitals, in the UK, via the contracting market process.

More examples of performance related funding can be found in the public funding mechanisms in the UK. The allocation of central grants to the local authorities is based on estimates of “efficient" costs. More recently, the performance related funding has emerged within the higher education where the allocation of resources seeks to take into account research and teaching performance of universities.

Critiques can be found in the literature (see for example the special issue of the journal of *Health Economics* (1994)) which concentrate on the limitations of the current efficiency assessment methods to estimate the "true" cost efficiency of public sector organisations. Moreover, problems of definition and measurement of quality (e.g. quality in the provision of health and measurement of research quality in the higher education sector) seem to cast doubts on how efficiency measures would operate in real decision making situations.

Relatively few voices can be found, however, calling for the combined use of the fundamental objectives of efficiency, equity and effectiveness within the legislative processes of allocating resources in not-for-profit organisations. In our opinion, there is scope for future research in this direction as the problem goes beyond the current critiques on matters of measurement and representation of appropriate input/output variables in the assessment of performance.

An attempt towards the simultaneous representation of efficiency-equity and effectiveness within an actual decision making framework for macroeconomic planning of the Greek local authorities is reported by Athanassopoulos (1994). Some very useful pointers have been obtained from this real life application insofar as the applicability of the models developed in chapter five is concerned. The project includes the participation of decision makers from the association of the Greek local authorities representing the local management (*agents*) and the ministers of interior and treasury in Greece representing the central management (*principals*) of the study. The pointers, listed below, are based on the use of planning models in line with the framework proposed in chapters three and five of this thesis. The pointers are summarised as follows:

- Different levels of decision making (e.g. central government and local authorities) have a very important and distinct role in the assessment of performance and the process of allocating resources.

- There are difficulties in defining commonly accepted input-output sets that describe the operations of multi-unit organisations (e.g. hospitals, schools, local authorities). This is the result of the different perceived objectives adopted by different levels of these organisations. This leads to the conclusion that

- The traditional assumption that decision making units are uniform can be questioned as there are cases where the decision making units are adopting different organisational missions which need to be reflected within control or planning models (e.g. local authorities, universities, schools).

- The viability of the process to reform a resource allocation system may rely on the power structure between those who gain and those who lose by the new system. Pressure groups (e.g. political parties), hence, can affect the attempts for reorganising the resource allocation process.

These pointers are put forward to highlight the significance and challenge of incorporating decision making preferences within a systematic process of resource allocation. The operationalisation of these systems can be pursued using group decision making and methods of negotiation that would facilitate the presence of qualitative aspects (e.g. human judgements) within mathematical decision making procedures.

4.1.1. *Implications for frontier analysis modelling*

The developments in the frontier analysis literature (see section 5 in Chapter 2) since its development by Farrell in (1957) have retained an efficiency assessment focus without particular concern on issues related with planning and decision making. The ability of frontier analysis to be used as a decision support tool will, in our opinion, be one of the determinants of the future viability of the method within the non-profit organisational arena. The original application of DEA in the public sector by Charnes et al (1981) sought to assess the degree of resource utilisation of different educational programmes in the USA. Thus, the message that frontier analysis should be linked with resource allocation and decision making has been indicated from the very early days of the methodology. The early experiments by Bessent et al (1982) and (1994) to support resource allocation decisions in the management of secondary education schools of San Antonio in Texas needs to be followed with methodological advancements and more applications.

The centralised and decentralised target based planning methods developed in chapters five and six put forward two alternative ways of addressing the problems of resource allocation and performance measurement in non-market oriented organisations. Undoubtedly, these models need to be extended in the future to take into account:

- Qualitative characteristics in the measurement of inputs/outputs,
- The dynamic character in the allocation of resources that have more than one fiscal year horizon,
- The uncertainty regarding quantities of inputs/outputs and also on the efficient performance of units over time.

The issues listed above would have profound implications regarding the use and application of the resource allocation methods of the thesis on existing organisations. The incorporation of qualitative aspects in the inputs/outputs used to describe operating processes is an issue that has not been addressed in full by the current frontier analysis literature, Olesen and Petersen (1993).

The extension of the resource allocation models developed in the thesis from a single to a multi-period problem can provide considerable support to the budgeting processes that cover, by nature, more than one time period. This extension will also bring forward strategic issues in the assessment of performance and the allocation of resources. The use of single period performance measurement and resource allocation models cannot capture strategic organisational issues. Thus, the development of multi-period resource allocation models is a necessary step towards enhancing the uses of the models concerned.

Finally, the incorporation of stochastic elements in resource allocation models is perhaps, a natural consequence of the previous proposal to introduce multi-period formulations. A considerable number of uncertainties are encountered in a resource allocation system: varying from the level of global resources the central management is prepared to allocate up to the very nature of the distribution of the inefficiencies within each firm. The work of Lovell et al (1993) could be considered here as the basis of building a stochastic programming framework for target based resource allocation. Furthermore, the use of stochastic factors within a dynamic framework of resource allocation can accommodate elements from the *principal-agent* theory. For example, Bogetoft (1994) has opened a prominent research area linking issues such as managerial effort and motivation with the improvement of efficiency.

4.2. Linking the diagnostic & planning tools developed in market based MUOs with decision making processes.

The *a-priori* decision support system was developed in the context of an existing organisation which indicated the need for some further research in this area. Such research should consider two main issues: improvements regarding the definition and assessment of market efficiency and the use of the information obtained by DEA to guide capital investment decision making.

4.2.1. *Market efficiency*

Market efficiency was introduced in the thesis as a measure of the extent to which individual retail outlets utilise their potential to attract custom. Evidently, the vast majority of efficiency assessment studies concentrate on the assessment of cost and not on market efficiency. The former has been characterised as an intrinsic factor of the performance of outlets and has widespread popularity in retail banking studies (see special issue of the Journal of Banking and Finance (1993), Volume 17).

The assessment of market efficiency is based on inputs/outputs that relate to the market conditions in the area surrounding an outlet. The definition, selection and measurement of this type of information is a crucial factor that affects the assessment of market efficiency. Factors such as competition and market potential, despite the difficulty in quantifying them have implications for the market efficiency models (see the discussion regarding the use of competition in DEA in chapter seven). Clearly, the representation of factors that determine the market conditions of individual outlets should be subject to systematic investigation prior to their use. DEA models assume causality between inputs and outputs in that the higher the input value is the higher should be the output value under efficient operation. Empirical studies see Mandley (1994) have shown that certain market conditions (e.g. number of competitors) may operate positively for some outputs and negative for some others. Modelling extensions will be necessary in order to address this type of problems. This is a direction that has been pursued by Athanassopoulos and Dyson (1994) where the efficiency is assessed without any assumptions about the positive or negative contribution of inputs on the generation of outputs during the assessment of efficiency.

4.2.2. From decision support to decision making

The justification of the *a-priori* modelling was focused on cases where individual DMUs are selected to undertake capital investment projects as a means of improved performance. The models developed in chapters seven, eight and nine had a decision support and not decision making character insofar as the direct allocation of capital is concerned. Future research should seek to explore the possibility of adding direct decision making character to those issues already developed in the thesis.

Some ideas towards that direction are next discussed. Decision making for improving performance of retail outlets is inextricably linked with the allocation of capital to undertake investment projects. As a further step of the current analysis one would expect to develop systematic models for linking the information provided in chapters seven to nine of this thesis with methods of investment appraisal. The mathematical programming literature provides sufficient background information to proceed towards these formulations. For example, multiple objective programming models for capital investment can be formulated (Thanassoulis (1983)) in order to help selecting projects that would :

i. Maximise expected returns on sales (market efficiency),

ii. Maximise expected profits,
iii. Maximise the long run viability of the network of DMUs,
iv. Minimise the projects' risk and uncertainty,
v. Maximise service mix effectiveness.

The information obtained from the analysis in chapters seven to nine can be used to represent quantitatively the objectives of capital investment appraisal. The objectives listed above are illustrative and can be modified appropriately for specific problems. The significance of capital investment problems is well appreciated within the literature. Yet, there are no sufficient answers concerning the estimation of expected returns from future investments and the identification of investment opportunities among networks of for-profit DMUs. It can be argued that the information yielded from the models in chapters seven to nine can enhance the effectiveness of the current investment appraisal methods.

This thesis sought to develop decision support mechanisms for improved resource management in multi-unit organisations. Two main cases were studied concerning *a-posteriori* and *a-priori* resource management problems. The model building process gave the opportunity to explore the potential of mathematical programming in general while frontier analysis in particular to be used in a *control* and *planning* mode. As Charnes and Cooper (1980) argue, mathematical programming has been traditionally used to support the solution of planning problems. The development of frontier analysis by Farrell (1957) and by Charnes *et al* (1978) gave a *control* dimension to mathematical programming. The current thesis attempted to contribute towards the integration the *control* and *planning* dimensions of mathematical programming as an aid to resource allocation and target setting.

In the years to come frontier analysis will be tested for its ability to expand its current horizons in a number of new directions. Decision support is one particular direction supported by this research. The ability of the research community to provide viable and realistic processes that would enable integration of the control and planning mechanisms in MUOs will affect the long run viability of frontier analysis.

- END OF THESIS -

REFERENCES

[1]. Achoff R., (1974), *Redesigning the Future: A systems approach to Societal problems.* New York: Wiley.

[2]. Adolphson D., G. Cornia and L. Walters, (1990), "A unified framework for classifying DEA models", in H. Bradley, ed., *Operational Research '90*, New York: Pergamon Press.

[3]. Aigner D., K. Lovell and P. Schmidt, (1977), "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics,* 6(1), pp. 21-37.

[4]. Aigner D. and S. Chu, (1968), "On estimating the industry production function", *American Economic Review* 58(4), pp. 826-839.

[5]. Ali A. and L. Seiford, (1989), "Computational accuracy and infinitesimals in data envelopment analysis", Technical report, University of Massachusetts at Amherst, Amherst.

[6]. Ali A. and L. Seiford, (1993), "The mathematical programming approach to efficiency analysis", In Fried H., K. Lovell, and S. Schmidt, *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York.

[7]. Allen D., (1994), "Targets, missions and criteria", *Management Accounting,* (January) pp.16.

[8]. Allison P., (1978), "Measures of inequality", *American Sociological Review,* Vol. 43, pp. 865-880.

[9]. Anandalingam G., (1988), "A mathematical programming model of decentralised multi-level systems", *Journal of the Operational Research Society,* Vol. 39, pp. 1021-1033.

[10]. Andersen P. and N. Petersen, (1994), "A procedure for ranking efficient units in data envelopment analysis", *Management Science,* Vol. 39, pp. 1261-1264.

[11]. Ansoff I., (1987), *Corporate strategy,* Penguin Business.

[12]. Antony R., (1965), *Planning and Control Systems: A framework for analysis,* Graduate School of Business Administration, Harvard University, Boston.

[13]. Argenti J., (1989), *Practical corporate planning,* Unwin.

[14]. Arrow K., (1963), "Uncertainty and the welfare economics of medical care", *American Economic Review,* LIII, (5), pp. 89-121.

[15]. Ashford J., M. Butts and T., Bailey, (1981), "Is there still a place for independent research into issues of public policy in England and Wales in the 1980 ?", *Journal of the Operational Research Society,* Vol. 32, pp. 851-864.

[16]. Ashton D., T. Hopper and R. Scapens, (1991), *Issues in management accounting*, Prentice Hall, London.

[17]. Athanassopoulos A. and E. Thanassoulis, (1995), "Separating market efficiency from profitability and its implications for planning", *Journal of Operational Research Society*, Vol. 46 (1), pp. 20-35.

[18]. Athanassopoulos A. and E. Thanassoulis, (Forthcoming 1995) "Assessing the marginal impacts of investment on the performance of organisational units", *International Journal of Production Economics.*

[19]. Athanassopoulos A. and N. Tatsos, (1992), "Econometric modelling as an aid for reorganisation policies: The case of the Greek local authorities", The Institute of Applied Econometrics, *Econometrics Europe 2000,* Belgium, Brussels, Published proceedings.

[20]. Athanassopoulos A. and J. Ballantine, (1993), "Ratio and frontier analysis for assessing corporate performance: Evidence from the Grocery industry in the UK", Warwick Business School research paper No. 101. (Forthcoming in the *Journal of the Operational Research Society, April 1995*).

[21]. Athanassopoulos A. and J. Storbeck, (1992), "Non-parametric models for Spatial Efficiency", (Forthcoming in the *Journal of Productivity Analysis).*

[22]. Athanassopoulos A. and R.G. Dyson, (1994), "Mathematical programming models for global resource allocation and target setting in multi-unit organisations", (Under review for publication in *Omega).*

[23]. Athanassopoulos A., (1994a), "Goal programming and Data Envelopment Analysis for target based multi-level planning: allocating central grants to Greek local authorities", Warwick Business School Research papers No. 144, Forthcomming in the *European Journal of Operational Research*).

[24]. Athanassopoulos A., (1994b), "Flexible frontier analysis models for target setting and decision support: an applciation to electricity generating plants", (Under revision for publication)

[25]. Atkinson A., (1970), "On the measurement of inequality", *Journal of Economic Theory*, Vol. 2, pp. 244-266.

[26]. Audit Commission, (1985), "Managing services effectively", Performance review No. 5.

[27]. Axelsson B. and G. Easton, (1992), "*Industrial networks: A new view of reality",* Routledge, London

[28]. Bailey A. and B. Warren, (1976), “Goal and resource transfers in the multi goal organisation”, *Accounting Review,* Vol. 51(3), pp. 559-573.

[29]. Banker R. and C. Kemerer. (1989). "Scale economies and new software development", *IEEE Transactions in Software Engineering,* 15(10), 1199-1205.

[30]. Banker R. and R. Thrall. (1992). "Estimation of returns to scale using Data Envelopment Analysis" *European Journal of Operations Research*, 62, 74-84.

[31]. Banker R. and R. Morey, (1993), "Integrated system design and operational decisions for service sector outlets", *Journal of Operations Management,* 11, 81-98.

[32]. Banker R., A. Charnes and W. Cooper, (1984), "Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis", *Management Science,* Vol.30(9), pp. 1078-1092.

[33]. Banker R., (1984), "Estimating most productive scale size using data envelopment analysis", *European Journal of Operational Research,* Vol. 17, pp. 35-44.

[34]. Banker R. and R. Morey, (1986a), "The use of categorical variables in data envelopment analysis", *Management Sciences*, Vol. 32, pp. 1613-1627.

[35]. Banker R. and R. Morey, (1986b), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research*, Vol. 34, pp.513-521.

[36]. Banker R., R. Conrad and R. Strauss, (1986), “A comparative application of data envelopment analysis and translog methods: An illustrative study of hospital production”, *Management Science,* Vol. 32, pp. 30-44.

[37]. Banker R. and R. Thrall, (1992), "Estima tion of returns to scale using data envelopment analysis", *European Journal of Operational Research,* Vol. 62, pp. 74-84.

[38]. Barrow, M. and Wagstaff, A., (1988), "Efficiency measurement in the public sector : An appraisal", *Fiscal Studies.*

[39]. Battese G. and G. Corra, (1977), "Estimation of a production frontier model: With application to the pastoral zone of eastern Australia", *Australian journal of Agricultural Economics,* 21(3), pp.169-179.

[40]. Baumol W. and R. Fabian, (1964), "Decomposition, pricing for decentralisation and external economies", *Management Science,* Vol. 11, pp. 23-37.

[41]. Beasley J., (1990), "Comparing university departments", *Omega International Journal of Management Science*, Vol. 18, No. 2, pp. 171-183.

[42]. Bennett R., (1980), *The geography of Public Finance: Welfare under fiscal federalism and local government finance,* Methuen, London.

[43]. Berg S., F. Forsund and E. Jansen, (1992), "Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-1992", *Scandinavian Journal of Economics,* Vol. 94, pp. 211-228.

[44]. Berne R. and L. Stiefel, (1984), *The measurement of equity in school finance: Conceptual, Methodological, and Empirical Dimensions,* Johns Hopkins University Press, Baltimore.

[45]. Bessent A., W. Bessent, J. Elam and T. Clark, (1988), "Efficiency frontier determination by constraint facet analysis", *Operations Research,* Vol. 36, pp. 785-796.

[46]. Bessent A., W. Bessent, J. Kennington and B. Reagan, (1982), "An application of mathematical programming to assess productivity in the Houston independent school district", *Management Science,* p.p. 1355- 1367.

[47]. Bessent A. and W. Bessent, (1984), "Educational productivity council employs management science methods to improve educational quality", *Interfaces*, 14, pp. 1-8.

[48]. Birch S. and A. Maynard, (1986), "Performance indicators and performance assessment in the UK national health service: Implications for management and planning", *International Journal of Public Health Planning,* Vol. 1, pp. 143-156.

[49]. Bodily S., (1992), "Introduction: The practice of Decision and Risk analysis", *Interfaces,* Vol. 22, pp. 1-4.

[50]. Bodily S., (1978), "Police sector design incorporating preferences of interest groups for equality and efficiency", *Management Science,* Vol. 24, pp. 1301-1312.

[51]. Bogetoft P., (1994), “Incentive efficient production frontiers: An agency percpective on DEA”, *Management Science,* Vol. 40, pp. 959-968.

[52]. Bolton P. and J. Farrell, (1990), "Decentralisation, duplication and delay", *Journal of Political Economy,* Vol. 98, pp. 803-826.

[53]. Bollen K., (1989), *Structural equations with latent variables,* New York: Wiley.

[54]. Booth P. and Cocks N., (1990), "Critical research issues in accounting standard setting", *Journal of Business Finance and Accounting,* pp. 511-528.

[55]. Bourgeois L., (1981), " On the measurement of organisational slack", *Academy of Management Review,* Vol. 6, pp. 29-39.

[56]. Boussofiane A., R. Dyson, and E. Thanassoulis, (1991), "Applied Data Envelopment Analysis" , *European Journal of Operations Research,* 52, pp.1-15.

[57]. Boussofiane A., R. Dyson and E. Thanassoulis, (1991), "Using data envelopment analysis to assess the efficiency of the perinatal care provision in England", Warwick Business School Research paper No. 5.

[58]. Bowlin W., J. Wallace and R., Murphy, (1989), Efficiency-based budgeting. *Journal of Cost Analysis*, Vol. 8, pp. 33-54.

[59]. Brooke M., (1984), *Centralisation and autonomy*, Chatham, Sussex

[60]. Brown C. and V. Jackson, (1990), *Public sector economics*, Oxford and Cambridge, Mass., Blackwell, 4th ed.

[61]. Brown F. and J. Pyers, (1988), "Putting teeth into the efficiency and effectiveness of public services", *Public Administration Review*, May/June, pp. 735-742.

[62]. Burns and Kaplan, (1987), "Accounting and Management", Harvard Business School Press.

[63]. Burton R. and B. Obel, (1977), "The multilevel approach to organisational issues of the firm - A critical review", *OMEGA The international Journal of Management Science*, Vol. 5, pp. 395-413.

[64]. Burton R. and B. Obel, (1988), "Opportunism, incentives and the M-form hypothesis", *Journal of Economic Behaviour and Organisation*, Vol. 10, pp. 99-119.

[65]. Butt H. and D. Palmer, (1985), *Value for Money in the Public sector*, Price Waterhouse, Basil Blackwell.

[66]. Butterworth P, R. Gray and J. Haslam, (1989), "The local authority annual report: An explanatory study of accounting communication and democracy", *Financial Accountability and Management*, 5(2), 73-88.

[67]. Byrne, Charnes, Cooper, *et al.* (1971), "Studies in Budgeting", North-Holland P.C.

[68]. Camp R., (1992), "Learning from the best leads to superior performance", *Journal of Business Strategy*, Vol. 13, pp.3-7.

[69]. Carr-Hill R., G. Hardman, S. Martin, S. Peacock, T. Sheldon and P. Smith, (1994), A formula for distributing NHS revenues based on small area use of hospital beds", Centre for Health Economics, The university of York.

[70]. Chambers R., (1990), *Applied production analysis; a dual approach"*, Cambridge University Press.

[71]. Chang K. and P. Guh, (1991), "Linear production functions and data envelopment analysis", *European journal of Operational Research*, 52(2), pp. 215-223.

[72]. Charnes A. and W. Cooper, (1959), "Chance-Constrained Programming", *Management Science*, 15, pp. 73-79.

[73]. Charnes A. and W. Cooper, (1978), "Managerial Economics Past, Present and Future", *Journal of Enterprise Management*, Vol. 1, pp. 5-23.

[74]. Charnes A., W. Cooper and E., Rhodes, (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research,* Vol. 2(6), pp. 429-444.

[75]. Charnes A, W. Cooper and E., Rhoades, (1979), "Short communication: Measuring efficiency of decision making units", *European Journal of Operational Research,* Vol. 3, pp. 339.

[76]. Charnes A., W. Cooper and E., Rhodes, (1981), "Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through", *Management Science,* Vol. 27, pp. 668-697.

[77]. Charnes A., W. Cooper, A. Lewin, R. Morey, and J. Rouseau, (1985a), "Sensitivity and stability analysis in DEA", *Annals of Operations Research,* Vol. 2, pp. 139-156.

[78]. Charnes A., C. Clark, W. Cooper and B. Golany, (1985b), "A development study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces", *Annals of Operations Research* 2, pp. 95-112.

[79]. Charnes A., W. Cooper, B. Golany, L. Seiford and J. Stutz, (1985c), "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions", *Journal of econometrics,* Vol. 30, pp. 91-107.

[80]. Charnes A. and W. Cooper, (1985d), "Preface to topics in Data Envelopment Analysis", *Annals of Operations Research,* Vol. 2, pp. 59-94.

[81]. Charnes A., W. Cooper and T. Sueyoshi, (1988), "A goal programming/constrained regression review of the Bell system break-up", *Management Science,* 34(1), pp. 1-26.

[82]. Charnes A., W. Cooper, Z. Huang and D. Sun, (1990), "Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks", *Journal of Econometrics,* Vol. 46, pp. 73-91.

[83]. Charnes A. and L. Neralic, (1990), "Sensitivity analysis of the additive model in data envelopment analysis", *European Journal of Operational Research,* Vol. 48, pp. 332-341.

[84]. Charnes A. and L. Neralic, (1991), "Sensitivity analysis of the simultaneous proportionate change of inputs and outputs in data envelopment analysis", Paper presented in the 11th European Congress on Operational Research, EURO-XI, Aachen, Germany.

[85]. Charnes A. and W. Cooper, (1977), " Goal programming and multiple objective optimisations", *European Journal of Operational Research,* Vol. 1, pp. 39-54.

[86]. Charnes A., W. Cooper, N. Eechambadi, B. Golany, D. Learner, F. Phillips and J. Rousseau, "A data envelopment analysis of high-turnover consumer products in

competitive markets", in A. Charnes, W. Cooper, and A. Lewin, eds., *New management application of data envelopment analysis,* to be published in 1993 by Kluwer Publishers.

[87]. Clarke R., (1992), "Evaluating USAF vehicle maintenance productivity over time: An application of Data Envelopment Analysis", *Decision Sciences*, Vol. 23, pp. 376-384.

[88]. Coates F., (1989), *Control and Audit in management accounting,* C.I.M.A.

[89]. Cook, W.D., A. Kazakov, Y. Roll and L. Seiford, (1991), "A Data Envelopment Approach to Measuring Efficiency: Case Analysis of Highway Maintenance Patrols", Journal of Socio-Economics, 20, pp. 83-103.

[90]. Cooper W. and A. Gallegos, (1991), "A combined DEA-stochastic frontier approach to latin American airline efficiency evaluations", Paper presented at the EURO XI Congress on Operational Research, Aachen, Germany.

[91]. Clarke R., (1992), "Evaluating USAF vehicle maintenance productivity over time: An application of Data Envelopment Analysis", *Decision Sciences*, Vol. 23, pp. 376-384.

[92]. Daft R., (1989), *Organisation Theory and Design,* West Publishing company, New York.

[93]. Danziger J., (1976), "Assessing the Incrementalism in British Municipal Budgeting", *British Journal of Political Science,* Vol. 6, pp. 335-350.

[94]. Danzing G. and P. Wolfe, (1961), "The decomposition algorithm for linear programming", *Econometrica,* Vol. 29, pp. 767-778.

[95]. Davis O., M. Dempster and A. Wildavsky, (1966), "A theory of the budgetary process", *American Political Science Review* 60: pp.529-547.

[96]. Day A. and A. Lewin, (Forthcomming 1995), "Strategic leaders in the U.S. Brewing industry: A longitudinal analysis of outliers, In: Charnes, Cooper, Lewin and Seiford, eds., *Data Envelopment analysis: Theory, methodology and applications.*

[97]. De Groot H., (1988), "Decentralisation decisions in bureaucracies as a principal-agent problem", *Journal of Public Economics,* Vol. 36, pp. 323-337.

[98]. Dearden J., (1969), "The case against R.O.I. control", *Harvard Business Review.*

[99]. Debreu G., (1951), "The coefficient of resource utilisation", *Econometrica,* 19(3), pp.273-292.

[100]. Deprins D., L. Simar and H. Tulkens, (1983), "Measuring labour efficiency of post offices", In Marchand, Petisteau and Tulkens, eds., *The Performance of public enterprises: concepts and measurement.* Amsterdam: North-Holland.

[101]. Dimsdale N. and N. Horsewood, (1992) "A model of the UK economy in the interwar period", *European Economic Review,* Vol. 36, pp. 702-709.

[102]. Drummond M., (1986), "Resource allocation decisions in health care: A role for quality of life assessments ?", *HSMC Discussion paper series* No.21.

[103]. Drury C., (1991), *Management and Cost accounting,* VNR Series in Accounting and Finance, 3rd edition, London.

[104]. Dyson R., (1990), *Strategic Planning Models and Analytical Techniques*, Wiley.

[105]. Dyson R., E. Thanassoulis and A. Athanassopoulos (1993), "Weights restrictions, value ratio frontiers, and production possibility frontiers in data envelopment analysis", Warwick Business School research papers No. 97.

[106]. Dyson R.G. and E. Thanassoulis, (1988), "Reducing weights flexibility in data envelopment analysis", *Journal of the Operational Research Society,* Vol. 39, pp. 563-576.

[107]. Eden C. (1988), "Cognitive mapping - A review", *European Journal of Operational Research,* Vol. 36, pp. 1-16.

[108]. Eisenhardt K. M., (1985), "Control: Organisational and Economic Approaches", *Management Science,* Vol. 31, No. 2, pp. 134-148.

[109]. Elyasiani E. and Mehidian S., (1990), "A nonparametric approach to measurement of efficiency and technological change: The case of large US. commercial banks", *Journal of Financial Services Research,* Vol. 4, pp. 157-168.

[110]. Enthoven A., (1985), "Reflections on the management of the National Health Service", Nuffield Provincial Hospitals Trust, London, Occasional papers No. 5.

[111]. Epstein M. and J. Henderson, (1988), "Data envelopment analysis for managerial control and diagnosis", *Decision Sciences,* 20(1), 90-119.

[112]. Etzioni A., (1964), *Modern Organisations,* Prentice-Hall.

[113]. Evanoff D. and P., Israilvevich, (1991), "Productive efficiency in banking", *Federal Reserve Bank Of Chicago.*

[114]. Evans D. and J. Heckman, (1988), "Natural monopoly and the Bell system: Response to Charnes, Cooper and Sueyoshi", *Management Science,* 34(1), pp. 27-38.

[115]. Fare R. and K. Lovell, (1978), "Measuring the technical efficiency of production", *Journal of Economic Theory,* 19(1), pp.150-162.

[116]. Fare R., S. Grosskopf, and C. Pasurka, (1983), "Effects on relative efficiency in electric power generation due to environmental controls", *Resources and Energy,* 8(2), 167-184.

[117]. Fare R., S. Grosskopf, and K. Lovell, (1985), *The measurement of efficiency of production,* Boston: Kluwer-Nijhoff publishing.

[118]. Fare R., (1988), *Fundamentals of Production theory,* Berlin: Springer-Verlang.

[119]. Fare R., S. Grosskopf, S. Yaisawarng, S. Li and S. Wang, (1990), "Productivity growth in Illinois electric utilities", *Resources and Energy,* Vol. 12, pp. 383-398.

[120]. Fare R., S. Grosskopf, B. Lindgren, P. Roos, (1992), "Productivity changes in Swedish pharmacies 1980-1989: A non-parametric Malmquist approach", *The Journal of Productivity Analysis,* Vol. 3, pp. 85-101.

[121]. Farrell M., (1957), "The measurement of productive efficiency", *Journal of Royal Statistical Society A,* Vol. 120, pp. 253-281.

[122]. Ferlie E., (1992), "The creation and evolution of quasi markets in the public sector: A problem for strategic management", *Strategic management journal,* 13, pp. 79-97.

[123]. Ferrier G. and K. Lovell. (1990). "Measuring cost efficiency in banking: Econometric and linear programming evidence", *Journal of Econometrics* 46(1), 229-245.

[124]. Field E. and M., Shutler, (1991), "Performance target ratios as a measure of management control", *OR-Insight,* Vol. 4, pp. 23-27.

[125]. Forsund F., (1992), "A comparison of parametric and non-parametric efficiency measures: The case of Norwegian ferries", *Journal of Productivity Analysis,* Vol. 3, pp. 25-43.

[126]. Forsund F. and L. Hjalmarsson, (1976), "Generalised Farrell measures of efficiency: an application to the milk processing in Swedish dairy plants", *The Economic Journal,* Vol. 80, pp. 294-315.

[127]. Forsund F. and L. Hjalmarsson, (1979), "Frontier production functions and technical progress: A study of general milk processing in Swedish dairy plants", *Econometrica,* 47(4), pp. 883-900.

[128]. Frantz R., (1992), "X-efficiency and allocative efficiency: what have we learned ?", *American Economic Review,* 82(2), pp. 434-438.

[129]. Freeland J., and J. Moore, (1977), "Implications of resource directive allocation models for organisational design", *Management Sciences,* Vol. 23, pp. 1050-1058.

[130]. Freeland J. and N. Baker, (1975), "Goal partitioning in a hierarchical organisation", *OMEGA the International Journal of Management Sciences*, Vol. 3, pp. 673-688.

[131]. Fried H., K. Lovell, and S. Schmidt, (1993), *The measurement of productive efficiency: Techniques and applications*, Oxford University Press, New York.

[132]. Galbraith J., (1977), *Designing Complex Organisations,* MacGraw-Hill, New York.

[133]. Gass S., (1986), " A process for determining priorities and weights for large scale linear programming models", *Journal of the Operational Research Society,* Vol. 37(8), pp. 779-785.

[134]. Gerberman, J., (1992), "On some problems of market planning evaluation and analysis in the presence of competitive and stochastic effects", *PhD-thesis,* University of Texas at Austin.

[135]. Ghosh A. and S. McLafferty, (1987), *Location strategies for retail and service firms,* Lexington Books, Toronto.

[136]. Giokas D., (1991), "Bank branch operating efficiency: A comparative application of DEA and the loglinear model", *Omega the International Journal of Management Science,* 19(6), 549-557.

[137]. Goicoechea A., D. Hansen and L. Duckstein, (1982), *Multiobjective Decision Analysis,* New Wiley.

[138]. Golany B., (1988), "An interactive MOLP procedure for the extension of DEA to effectiveness analysis", *Journal of the Operational Research Society,* Vol. 39(8), pp. 725-734.

[139]. Golany B., F. Phillips and J. Rousseau. (1991). "Models for improved effectiveness based on DEA efficiency results", *IIE Institute of Industrial Engineering Transactions.*

[140]. Golany B. and Roll. (1993). "Some extensions of techniques to handle non-discretionary factors in data envelopment analysis", *Journal of Productivity Analysis,* 4, 419-432.

[141]. Gold B., (1981), "Changing perspectives on size, scale and returns: an interpretive survey", *Journal of Economic Literature* **XIX**, 5-33.

[142]. Gold B., (1985), "Productivity measures: Descriptive averages versus analytical needs", In. *Managerial Issues in Productivity Analysis,* Eds. A., Dogramaci and N., Adam, Kluwer-Nijhopff.

[143]. Govindarajan V., (1986), "Impact of participation in the budgetary process on managerial attitudes and performance: Universalistic and contingency perspectives", *Decision Sciences,* Vol. 17, pp. 496-516.

[144]. Griffiths I., (1992), *Creative accounting: How to make your profits what you want them to be",* Routledgw, London.

[145]. Grosskopf S. and S. Yaisawarng, (1990), "Economies of Scope in the provision of local public services", *National Tax Journal,* Vol. 43, pp. 61-74.

[146]. Hadley G., (1961), *Linear Programming,* Narosa Publishing house, New Delhi.

[147]. Hannan E., (1981), "An assessment of some of the criticisms of goal programming, *Computers and Operations Research,* Vol. 12, pp. 525-541.

[148]. Hancock K., D. Holden, K. Swales, (1991), "Consistency of the spatial allocation of NHS hospital resources: An econometric analysis, Vol. 23, pp. 1623-1636.

[149]. Hax, A., and N., Majluf, (1983), " The use of the growth-share matrix in strategic planning", *Interfaces,* Vol. 13, No. 1, pp. 46-60.

[150]. Heiner K., W. Wallace and K. Young, (1981), "A resource allocation and evaluation model for providing services to the mentally retarded", *Management Science,* Vol. 27, pp. 769-784.

[151]. Henderson, B., (1970), "The product portfolio", The Boston Consulting Group, Perspectives No. 66, Boston, Massachusetts.

[152]. Hihn J. and C. Johnson, (1988), "Evaluation Techniques for Paired Ratio-Comparison in a Hierarchical Decision model", *Measurement in Economics,* eds. W., Eichhorn, Physica-Verlang Heidelberg.

[153]. Hoagland J., (1964), "Historical antecedents of organisational research, In: Cooper, Levvit and Shelly: *New perspectives in organisational research,* John Wiley & Sons.

[154]. Hogler R., and H. Hunt III, (1993), "Accounting and conceptions of control in the American corporation", *British Journal of Management,* Vol. 4, pp. 177-190.

[155]. Hwang C., and A. Massud, (1979), *Multiple objective decision making- Methods and applications,* Lecture notes in Economics and Mathematical Systems, Springer-Verlag, New York.

[156]. Ignizio J., (1982), *Linear programming in single and multiple objectives systems,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[157]. Johansen L., (1978), *Lectures in microeconomic planning,* North Holland, Amsterdam.

[158]. Jesson, D., D., Mayston and P. Smith, (1987), "Performance assessment in the educational sector: educational and economic perspectives", *Oxford review of education,* Vol. 13, No. 3, pp. 249-265.

[159]. Johnson C., and K. Scholes, (1988), *Exploring corporate strategy,* Prentice Hall.

[160]. Jondrow J., K. Lovell, S. Materov, and P. Schmidt, (1982), "On the estimation of technical inefficiency in the stochastic frontier production function model", *Journal of econometrics* 19(2/3), pp. 233-238.

[161]. Kaplan, R., and Atkinson A., (1989), "*Advanced management Accounting*", Prentice-Hall International editions

[162]. Kaplan R., (1983), "The evolution of management accounting", *The Accounting review,* Vol. LVIII, No. 4, pp. 686-705.

[163]. Kaplan R., (1982), *Advanced Management Accounting,* Prentice-Hall, Englewood Cliffs NJ.

[164]. Keeney R., and H. Raiffa, (1976), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs,* Wiley.

[165]. Khorramshahgol R., and V. Moustakis, (1988), "Delphi hierarchy process (DHP): A methodology for priority setting derived from the Delphi method and analytical hierarchy process", *European Journal of operational Research,* 37, pp. 347-354.

[166]. Kittelsen, S., (1993), "Stepwise DEA; choosing variables for measuring technical efficiency in Norwegian electricity distribution", Paper presented at the *Third European Workshop on Efficiency and Productivity Measurement,* Universite Catholique de Louvain.

[167]. Knight F., (1933), *The economic organisation,* New York: Harper & Row.

[168]. Koopmans (1951) *Activity analysis of production and allocation,* John Wiley and Sons, New York.

[169]. Koopmans (1957), *Three essays on the state of economic science,* McGraw-Hill, New York.

[170]. Kornai J., (1973), “Thoughts on multi-level planning systems”, In L. Coreux and A. Manne, eds., *Multi-level planning: case studies in Mexico,* North-Holland.

[171]. Kornbluth J. and G. Salkin, (1987), *The management of corporate financial assets: applications of mathematical programming models,* Academic Press, London.

[172]. Korostelev A., L. Simar and A., Tsybakov, (1993), "Efficient estimation of monotone boundaries", Paper presented at the *Third European Workshop on Efficiency and Productivity Measurement,* Universite Catholique de Louvain.

[173]. Kotler, (1989), *Marketing management*, Prentice-Hall International editions.

[174]. Land K., K. Lovell and S. Thore, (1991), "Change-constraint data envelopment analysis", *IC^2 Institute discussion papers*, The university of Texas at Austin.

[175]. Land K., K. Lovell and S. Thore, (1993), "Efficiency under capitalism and state socialism", *Public Finance - Finances Publiques,* 47, pp. 109-121.

[176]. Lanen W. and D. Larcker, (1992), "Executive compensation contract adoption in the electric utility industry", *Journal of Accounting Research,* Vol. 30, pp. 70-93.

[177]. Lapsley, I., (1988), "Capital Budgeting, Public Service organisations and U.K. Governmental Policy, *Journal of Accounting and Public Policy,* 7, pp. 65-74.

[178]. Leibenstein H., (1976), *Beyond Economic Man,* Cambridge, Mass.: Harvard University Press.

[179]. Leibenstein H., (1987), *Inside the firm,* Cambridge, Mass.: Harvard University Press.

[180]. Leibenstein H. and S. Maital, (1992), "Empirical estimation and partitioning of X-efficiency: A data envelopment approach", *American Economic Review,* Vol. 82, pp 428-433.

[181]. Lerne K., (1992), "Efficiency and frontier analysis with extension to strategic planning", *PhD-thesis,* University of Massachusetts at Amherst.

[182]. Lewin A. and D. Morey (1981), "Measuring the relative efficiency and output potential of Public sector organisations: An application of DEA", *International Journal Of Policy Analysis and Information Systems,* Vol 5, pp. 267-285.

[183]. Lewin A. and J. Minton, (1986), "Determining organisational effectiveness: Another look, and an agenda for research", *Management Science,* 32(5), pp. 514-537.

[184]. Lindblom,C., (1959), "The Science of "Muddling Through", *Public Administration Review,* 19, pp. 79-88.

[185]. Lineberry R., (1977), *Equality and urban policy,* Sage publications, California.

[186]. Lovell K. and R. Morey, (1991), "The allocation of consumer incentives to meet simultaneous sales quotas: An application to U.S. Army recruiting, *Management Science,* Vol. 37, No 3, pp. 350-367.

[187]. Lovell K., (1993), "Production frontiers and productive efficiency", In H. Fried, K. Lovell, and S. Schmidt, eds., *The measurement of productive efficiency techniques and applications,* Oxford University Press.

[188]. Lovell K., L. Walters and L. Wood, "Stratified models of education production using DEA and regression analysis", in A. Charnes, W. Cooper, A. Lewin and L. Seiford, eds. *Data envelopment analysis: Theory, Method and Process,* IC^2 Management and Management Science series, New York: Quorum Books, Forthcoming (1995).

[189]. Manzini A. and P. Thalmann, (1994), "The dynamic allocative efficiency of a public utility: Swiss Telecomunications", *Swiss Journal of Economics and Statistics,* Vol. 130, pp. 129-144.

[190]. Mackenzie, A., (1990), "Assessing the performance of Cadbury's sales forces" Unpublished MSc thesis, Warwick Business School

[191]. Mahajan J., (1991), "A data envelopment analysis model for assessing the relative efficiency of the selling function", *European Journal of Operational Research,* 53(2), 189-205.

[192]. Manne A., (1973), "Economic alternatives for Mexico", In L. Coreux and A. Manne, eds., *Multi-level planning: case studies in Mexico,* North-Holland.

[193]. Mandell, M., (1991), "Modelling effectiveness-Equity tradeoffs in public service delivery systems", *Management Science,* Vol. 37, pp. 467-482.

[194]. Mandley A., (1994), *A study into the use of DEA for assessing relative branch efficiencies in a retail banking network,* Unpublished Msc thesis, University of Warwick.

[195]. Marakon Associates, (1980), "Criteria for determining an optimal business portfolio", Presentation made at the Institute of Management Sciences, November 11.

[196]. March M., and D. Schilling, (1994), "Equity measurement in facility location analysis: A review and framework", *European Journal of Operational Research,* Vol. 74, pp. 1-17.

[197]. Mayston D., (1985), "Non-profit performance indicators in the Public Sector", *Financial Accountability and Management,* Vol. 1, pp. 51-74.

[198]. Mayston D., (1993), "Principals, agents and the economics of accountability in the new public sector", *Accounting Auditing and Accountability Journal,* 6(3), pp. 68-96.

[199]. McCarty T. and S. Yaisawarng, (1993), "Technical efficiency in New Jersey school districts", In H. Fried, K. Lovell, and S. Schmidt, eds., *The measurement of productive efficiency techniques and applications,* Oxford University Press.

[200]. Metzger L., (1992), "Measuring quality cost effects on productivity using data envelopment analysis", *Journal of Applied Business Research,* 9(3), pp. 69-79.

[201]. Min H. and J. Storbeck, (1991), "On the origin and persistence of misconceptions in goal programming", *Journal of the Operational Research Society,* Vol. 42, pp. 301-312.

[202]. Monk D., (1986), "Social grading and the national readership survey", *Joint Industrial Committee for National Readership Survey* (JICNARS).

[203]. Moore H., (1929), *Synthetic economics* , New York, Macmillan.

[204]. Nàchane D., (1984), "Optimisation methods in multilevel systems: A methodological survey", *European Journal of Operational Research,* Vol. 21, pp. 25-38.

[205]. National Audit Office ((NAO0 1992), *Report of controller and auditor general on department of health appropriation accounts,* HMSO, London.

[206]. Nijkamp P. and P. Rietveld, (1981), "Multi-objective multi-level policy models: An application to regional and environmental planning", *European Economic Review,* Vol. 15, pp. 63-89.

[207]. Novick D., (1973), *Current practice in Program Budgeting (PPBS),* Heinemann, London.

[208]. Olesen O. and N. Petersen, (1991), "Collinearity in data envelopment analysis: An extended facet approach", Working paper No. 1/1991, Department of management, Odense University, Denmark.

[209]. Olesen O. and N. Petersen, (1993), "Incorporating quality into data envelopment analysis: A stochastic dominance approach", Working paper No. 2/1993, Department of management, Odense University, Denmark.

[210]. Olson D. and V. Dorai, (1992), "Implementation of the centroid method of Solymosi and Dombi", *European Journal of Operational Research,* Vol. 60, pp. 117-129.

[211]. Ouchi W., (1979), "A conceptual framework for the design of organisational control mechanisms", *Management Science,* Vol. 25, pp. 833-847.

[212]. Perrow C., (1961), "The analysis of goals in complex organisations", *American sociological review,* 26, pp. 854-866.

[213]. Peters T., (1992), *Liberation Management,* Macmillan, London.

[214]. Petersen N., (1990), "Data envelopment analysis on a relaxed set of assumptions", *Management Science*, 36(3), pp. 305-314.

[215]. Petisteau P. and H. Tulkens, (1990), "Assessing the performance of public sector activities: Some recent evidence from the productive efficiency viewpoint", Discussion paper No. 9060, CORE, Louvain-la-Neuve, Belgium.

[216]. Preston A., (1991), "Budgeting, creativity and culture", In *Issues in management accounting,* D. Ashton, T. Hopper, R. Scapens, eds., Prentice-Hall, New York.

[217]. Quinn R. and J., Rohrbaugh (1981), "A competing values approach to organisational effectiveness", *Public Productivity Review,* Vol. 5, pp. 139.

[218]. Ray S., (1988), "Data envelopment analysis, non-discretionary inputs and efficiency: An alternative interpretation", *Socio-Economic Planning Sciences,* 22(4), pp. 167-176.

[219]. Richards M., (1986), *Setting strategic goals and objectives,* The West series in strategic management, West Publishing company, New York.

[220]. Rockatt J., (1979), "Chief executives define their own data needs", *Harvard Business Review.*

[221]. Romero C., (1991), *"Handbook of critical issues in goal programming",* Pergamon press, Oxford.

[222]. Rotch, W. (1993). Management control systems: One view of components and their interdependence. *British Journal of Management,* Vol. 4, pp. 177-193.

[223]. Roy B., (1993), "Decision science or decision-aid science", *European Journal of the Operational Research,* Vol. 66, pp. 184-203.

[224]. Ruefli T., (1971), "PPBS-an analytical approach", In *Studies in Budgeting,* Edited by R. Byrne, pp. 161-261, North-Holland, Amsterdam.

[225]. Ruefli T., (1974), "Analytical models of resource allocation in hierarchical multi-level systems", *Socio-Economic Planning Sciences,* Vol 8, pp. 353-363.

[226]. Russell R., (1985), "Measures of technical efficiency", *Journal of economic theory,* 35(1), pp. 109-126.

[227]. Saaty L., (1980), *The analytic Hierarchy Process,* McGraw-Hill, New York

[228]. Savas E., (1978), "On equity in providing public services", *Management Science,* Vol. 24, pp. 800-808.

[229]. Schick A., (1990), "Budgeting for Results: Recent Developments in Five Industrialised Countries", *Public Administration Review, January/February, pp.26-34.*

[230]. Schmidt P., (1985-1986), "Frontier production functions", *Econometric reviews,* 4(2), pp. 289-328.

[231]. Schoeffler S., R. Buzzell and D., Heany, (1974), "Impact of strategic planning on profit performance", *Harvard Business Review,* March/April, pp. 137-145.

[232]. Seiford L., (1990), *A bibliography of Data Envelopment Analysis (1978-1990).* Technical report, Department of Industrial Engineering and Operations Research, University of Massachusetts, Amherst.

[233]. Seiford L. and R. Thrall, (1990), "Recent developments in Data Envelopment Analysis", *Journal of Econometrics,* Vol. 46, pp. 7-38.

[234]. Sexton T., A. Leiken, A. Nolan, S. Liss, A. Hogan, and S. Silkman, (1991), "Data envelopment analysis for non-homogeneous units: Pupils transportation budgeting in North Carolina", Working paper, The State University of New York, Stony Brook, N.Y.

[235]. Shephard R., (1953), *Cost and Production functions*, Princeton, Princeton University Press.

[236]. Sherman D. and B. Gold, (1985), "Bank Branch Operating Efficiency : Evaluation with Data Envelopment Analysis" *Journal of Banking and Finance*, Vol 9, No 2, pp. 297-315.

[237]. Sherman D., (1981), "Measurement of hospital technical efficiency: A comparative evaluation of data envelopment analysis and other approaches for locating inefficiencies in health care organisations" *DBA thesis, Harvard University.*

[238]. Sherman D., (1989), "*Service organisation productivity measurement*", The society of management accountants of Canada.

[239]. Sherman D., (1984), "Data envelopment analysis as a new Managerial Audit methodology - Test and Evaluation", *Auditing: A journal of practice and theory*, Vol 4, No.1, pp. 35 -53.

[240]. Sinuany-Stern Z., (1984), "A network optimisation model for budget allocation in a multi-campus university", *Journal of the Operational Research Society*, Vol. 35, pp. 749-757.

[241]. Smith J., (1993), "Modelling muddles: Validation beyond the numbers", *European Journal of Operational Research*, Vol. 66, pp. 235-249.

[242]. Smith P., (1987), "Optimal local authority budgeting strategies under block grant", *Applied Economics*, Vol. 19, pp. 891-905.

[243]. Smith P., (1990), "The use of performance indicators in the public sector", *Journal of Royal Statistical Society*, Series A, Vol. 153, pp. 53-72.

[244]. Smith P., (1992), "Negative political feedback: An examination of the problem of modelling political responses in public sector effectiveness auditing", *Accounting, Auditing and Accountability Journal*, Vol. 5, pp. 5-20.

[245]. Smith P., (1993), "Outcome-related performance indicators and organisational control in the public sector", *British Journal of Management*, Vol. 3, pp. 135-151.

[246]. Smith P. and D. Mayston, (1987), "Measuring efficiency in the public sector", *Omega*, Vol. 15, pp. 181-189.

[247]. Smith T., (1992), *Accounting for growth: Stripping the camouflage from company accounts*, Century, London.

[248]. Solomons D., (1978), "The politicisation of accounting", *The Journal of Accountancy*, November, pp. 65-72.

[249]. Srinivasan V., (1973), "Estimating the weights for multiple attributes in a composite criterion using pairwise judgements", *Psychometrika*, 38, pp. 473-493.

[250]. Steuer V. (1983), *Multiple criteria optimisation: Theory, computation and application,* John Wiley, New York.

[251]. Stigler G., (1976), "The existence of X-efficiency", *American Economic Review,* Vol. 66, pp. 213-216.

[252]. Storbeck J., (1992). "Decision support for retailing networks", Paper presented in the *National Operational Research* conference, Birmingham.

[253]. Sweeney D., P. Winkofsky, P. Roy and N. Baker, (1978), "Composition Vs. decomposition: Two approaches to modelling organisational decision processes", *Management Science,* Vol. 24, pp. 1491-1498.

[254]. Taylor F., (1911), *Principles of Scientific Management,* Harper and Bros., New York.

[255]. Taylor S., (1989), "Measuring need to spend on education", *Journal of the Operational Research Society,* Vol. 40, pp. 177-186.

[256]. Ten Kate A., (1972), "Decomposition of linear programs by direct distribution", *Econometrica,* Vol. 40, pp. 883-898.

[257]. Thanassoulis E., (1985), "Selecting a suitable solution method for a multi objective programming capital budgeting problem", *Journal of Business Finance and Accounting,* Vol. 12, pp. 453-471.

[258]. Thanassoulis E. and R.G. Dyson, (1992), "Estimating preferred targets input-output levels using Data Envelopment Analysis", *European Journal of Operational Research,* Vol. 56, pp. 80-97.

[259]. Thanassoulis E., A. Boussofianie and R.G. Dyson, (1995), "Exploring output quality targets in the provision f perinatal care in England using DEA", Forthcoming in the *European Journal of Operational Research.*

[260]. Thanassoulis E., (1993), "A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments", *Journal of the Operational Research Society,* 44(11), 1129-1145.

[261]. Thanassoulis E. and P. Dunstand, (1993), "Guiding schools to improved performance using data envelopment analysis: An illustration with data from a Local Education Authority", Warwick Business School research papers No. 94, (Forthcoming in the *Journal of the Operational Research Society*)

[262]. Thanassoulis E. and R.G. Dyson, (1992), "Estimating preferred input/output levels using data envelopment analysis", *European Journal of Operations Research,* Vol. 56, pp. 88-97. (In the Warwick Business School research papers in (1988)).

[263]. The Warwick-DEA package (user guide), (1992), Warwick Business School, University of Warwick.

[264]. Thiry B. and H. Tulkens, (1992), "Allowing for technical inefficiency in parametric estimation of production functions for urban transit firms", *Journal of Productivity Analysis,* Vol. 3, pp. 45-65.

[265]. Thompson G., F. Singleton, R. Thrall and B. Smith, (1986), "Comparative Site Evaluations for Locating a High-Energy Physics Lab In Texas", *Interfaces,* Vol.16(6), pp. 35-49.

[266]. Thompson R., L. Langemeier, C. Lee, E. Lee and M., Thrall, (1990) "The role of multiplier bounds in efficiency analysis with application to Kansas farming", *Journal of Econometrics,* Vol. 46, pp. 73-93.

[267]. Thurow L., (1976), "Equity versus efficiency in Law enforcement", in *The Economic Approach to Public Policy,* R. Amacher, R., Tollison, and Willett, eds. Cornell University Press, Ithaca, New York, pp. 124-131.

[268]. Tingley K. and J. Liebman, (1984), "A goal programming example in public health resource allocation", *Management Science,* Vol. 30, pp. 279-289.

[269]. Tulkens H., P. Vanden Eeckaut and M., Jamar, (1993), "Cost efficiency in Belgian municipalities", In H. Fried, K. Lovell and S. Schmidt, eds., *The measurement of productive efficiency:Techniques and applications,* Oxford University Press.

[270]. Tulkens H. and P. Vanden Eeckaut, (1991), "Non-frontier measures of efficiency, Progress and Regress", CORE discussion paper 9155, *Centre of Operational Research and Econometrics,* Universite Catholique de Louvain.

[271]. Van Ackere A., (1993), "The principal / agent paradigm: Its relevance to various functional fields", *European Journal of Operational Research,* Vol. 70, pp. 83-103.

[272]. Van De Panne C., (1991), "Decentralisation for multidivision enterprises", *Operations Research,* Vol. 39, pp. 786-797.

[273]. Varian H., (1984), *Microeconomic Analysis,* W.W. Norton & Company, London.

[274]. Vassiloglou M. and D. Giokas, (1990), "A study of the relative efficiency of bank branches: An application of data envelopment analysis", *Journal of the Operational Research Society,* 41(7), 591-597.

[275]. Wilson,D., and R., Rosenfeld, (1990), *Managing organisations,* McGraw-Hill, London.

[276]. Yue, P., (1992), "Data Envelopment Analysis and Commercial bank performance: A primer with applications to Missouri banks", *Federal Reserve Bank of St. Liouis,* pp. 31-45.

[277]. Zanakis S. and S. Gupta, (1985), "A categorised bibliographic survey of goal programming", *Omega* , Vol. 13, pp. 211-222.

[278]. Zeleny M., (1982), "The pros and cons of goal programming", *Computers and Operations Research,* Vol. 10, pp. 357-359.

[279]. Zeleny M., (1982), *Multiple criteria decision making,* McGraw-Hill, New York.

[280]. Zeleny M., (1980), "Multiple objectives in mathematical programming : Letting the Man in", *Computers & Operational Research*, Vol. 7, pp. 1-4.

[281]. Zieschang K., (1984), "An extended Farrell technical efficiency measure", *Journal of Economic Theory,* Vol. 33, pp. 387-396.

[282]. Zieschang K., (1983), "A note on the decomposition of cost efficiency into technical and allocative components", *Journal of Econometrics,* Vol. 23, pp. 401-405.

[283]. Zionts S. and J. Wallenius, (1976), "An interactive programming method for solving the multiple criteria problem", *Management Science,* Vol. 22, pp. 652-663.

APPENDICES OF CHAPTER 4

APPENDIX 4A

THE DUAL FORMULATION OF THE FLEXIBLE PRIORITISED MODEL

The flexible prioritised model in M4.3 can be written in a dual form similar to the one developed for the basic prioritised model in M4.6. This formulation is given in M4A.1

$$\min_{\alpha_r,\beta_i,\gamma_r^{\Gamma},\gamma_r^{U},\delta_i^{\Delta},\delta_i^{U}} -\sum_{r\in O}\gamma_r^{\Gamma}\Gamma_r+\sum_{r\in O}\gamma_r^{U}\, 1/U_r-\sum_{i\in I}\delta_i^{L}L_i+\sum_{i\in I}\delta_i^{\Delta}\, 1/\Delta_i \qquad \text{(M4A.1)}$$

$$\sum_{i\in I}\beta_i x_{ij}-\sum_{r\in O}\alpha_r y_{rj}\geq 0 \qquad \forall j \qquad (4A.1a)$$

$$-\beta_i x_{ij_o}-\delta_i^{L}+\delta_i^{\Delta}\geq -P_i \qquad \forall i \qquad (4A.1b)$$

$$\alpha_r y_{rj_o}-\gamma_r^{\Gamma}+\gamma_r^{U}\geq P_r \qquad \forall r \qquad (4A.1c)$$

$$L_i,\Gamma_r,\Delta_i,U_r\in[0,1]$$

$$\delta_i^{L},\delta_i^{\Delta},\gamma_r^{\Delta},\gamma_r^{U}\geq 0$$

$$\alpha_r,\beta_i \text{ free}.$$

The notation is similar to the one used in the primal formulation in M4.3 with,

x_{ij}, y_{rj}	are the ith input and rth output of DMU j,
P_i, P_r	are user specified priorities over input/output improvements,
Δ_i, Γ_r	are lower bounds of desired improvement and controllability of inputs and outputs,
L_i, U_r	are upper bounds of desired improvement and controllability of inputs and outputs,
α_r, β_i	are weight variables for output r and input i respectively,
$\gamma_r^{\Gamma,\Delta}, \delta_i^{\Gamma,\Delta}$	are weight variables for the upper/lower bounds of the rates of improvement of output r and input i respectively.

The set of constraints in M4A.1a represents the condition that the weighted sum of outputs of assessed units should not exceed the corresponding weighted sum of inputs. This is the condition upon which all known DEA models are built. The remaining constraints M4A.1b and M4A.1c seek to incorporate lower bounds on the contribution of individual inputs/outputs on the weighted sum of inputs or outputs of individual units.

APPENDIX 4B

TARGETS ESTIMATED BY THE FLEXIBLE TARGETS MODEL IN M4.3 ARE EFFICIENT

The proof of the theorem is very similar to the ones employed by Thanassoulis and Dyson (1992) to prove the estimation of efficient targets by the prioritised targets model.

The prioritised models is reproduced below:

$$
\begin{array}{llll}
\underset{\lambda_j, z_r, \theta_i}{Max} & \sum_{r\in O} P_r^+ z_r - \sum_{i\in I} P_i^- \theta_i & & \\
s.t. & \sum_{j=1}^{n} \lambda_j x_{ij} = \theta_i x_{ij_o} & \forall i \in I & \\
& \sum_{j=1}^{n} \lambda_j y_{rj} = z_r y_{rj_o} & \forall r \in O & \text{(PT)} \\
& L_i \le \theta_i \le 1/\Delta_i, \quad \Delta_i, L_i \in [0,1] & \forall i \in I & \\
& U_r \le 1/z_r \le 1/\Gamma_r, \quad \Gamma_r, U_r \in [0,1] & \forall r \in O & \\
& \lambda_j \ge 0 \quad \forall j & &
\end{array}
$$

Let us assume that a DMU j_O has been assessed using the PT model and its was found relatively inefficient, i.e. $\exists r : z_r \neq 1$ *or* $\exists i : \theta_i \neq 1$. The targets $(\hat{y}_{rj_o}, \hat{x}_{ij_o})$ estimated for unit j_O are obtained as follows: $\hat{y}_{rj_o} = z_r^* y_{rj_o}$ and $\hat{x}_{ij_o} = \theta_i^* x_{ij_o}$.

If we adjust the input/output levels of unit j_O to its target levels $(\hat{y}_{rj_o}\ \hat{x}_{ij_o})$ the PT can be employed to test whether these input/output levels make j_O efficient. The revised PT1 model is given below:

$$
\begin{array}{llll}
\underset{\tau_j, o_r, \iota_i}{Max} & \sum_{r\in O} P_r^+ o_r - \sum_{i\in I} P_i^- \iota_i & & \\
& \sum_{j=1, j\neq j_o}^{n} \tau_j x_{ij} + \tau_j \hat{x}_{ij_o} = \iota_i \hat{x}_{ij_o} & \forall i \in I & \\
& \sum_{j=1, j\neq j_o}^{n} \tau_j y_{rj} + \tau_j \hat{y}_{rj_o} = o_r \hat{y}_{rj_o} & \forall r \in O & \text{(PT1)} \\
& L_i \le \iota_i \le 1/\Delta_i, \quad \Delta_i, L_i \in [0,1] & \forall i \in I & \\
& U_r \le 1/o_r \le 1/\Gamma_r, \quad \Gamma_r, U_r \in [0,1] & \forall r \in O &
\end{array}
$$

If the optimal solution to the PT1 model is $\{o_r^*, \iota_i^*, \tau_j^* \mid \exists r : o_r^* \neq 1 \text{ or } \exists i : \iota_i^* \neq 1\}$ then this implies that unit j_O is inefficient.

The optimal solution to the PT1 problem $\left(\iota_i^* \hat{x}_{ij_o}, o_r^* \hat{y}_{rj_o}\right)$ can be used, therefore, as a feasible solution to the original PT problem. The optimal solution obtained from PT1 is a feasible solution of PT since the target input-output level $\left(\iota_i^* \hat{x}_{ij_o}, o_r^* \hat{y}_{rj_o}\right)$ is a linear combination of the original input-output set $\left(x_{ij_o}, y_{rj_o}\right)$.

This would yield a new feasible solution $\{z_r^f, \theta_i^f, \lambda_j^f\}$ as follows:

$$\begin{aligned} z_r^f &= z_r^* \times o_r^* \\ \theta_i^f &= \theta_i^* \times \iota_i^* \\ \lambda_j^f &= \tau_j^* \end{aligned} \qquad \text{(a)}$$

Given the set of preferences P_r^+, P_i^-, the feasible targets obtained in (a) dominate the optimal solution for unit j_O obtained from the solution to the original PT problem. This is not possible as the original solution to PT is optimal and therefore we conclude that the targets estimated by the PT model are efficient as they cannot be improved any further without violating optimality conditions.

APPENDIX 4C

POLICY MAKING CONDITIONS IN THE FLEXIBLE TARGETS MODEL

The flexible target setting model introduced in the main part of the thesis (M4.3) seeks to assess input/output targets constrained within policy making upper and lower bounds. Some extensions concerning the direction of the assessed targets are introduced below.

The simplest way to monitor the directions of the assessed targets is to use the weights of preferences included in the objective function of the flexible targets' model. If $P_r > P_i$ for all outputs/inputs in the analysis then a priority is given to the model for exploring output expansion targets as opposed to input contraction ones.

A more systematic process need to the used, however, in case the decision makers need to link the expansion/contraction of selected inputs with the expansion contraction of selected outputs. We illustrate next two possible cases:

If we assume that $z_r < 1$ then the following is expected $z_r > \theta_i$ for some selected r, i

If we assume that $\theta_i > 1$then the following is expected $z_r > \theta_i$ for some selected r, i

The condition above seeks to guarantee that if a unit reduces its produced output then it should also reduce its level of input use to more than compensate for the output reduction. On the other hand, if the unit increases its resource availability is should increase the production of outputs to compensate for the cost of the extra inputs. The mathematical formulation of these conditions requires the use of zero-one control variables and it is provided below:

$$\underset{\theta_i, z_r, \lambda_j}{Max} \quad \sum_{r \in O} P_r^+ z_r - \sum_{i \in I} P_i^- z_i$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} \qquad = \theta_i x_{ij_o} \qquad i \in I$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \qquad = z_r y_{rj_o} \qquad r \in O$$

$$l_i \leq \ \theta_i \ \leq 1/u_i, \quad u_i, l_i \in [0,1], \quad \forall i \in I$$

$$u_r \leq 1/z_r \leq 1/l_r, \quad l_r, u_r \in [0,1], \quad \forall r \in O$$

$$\begin{gathered} -(z_r - \theta_i) \leq M\delta \\ (-z_r + 1) \leq M(1-\delta) \\ -(z_r - \theta_i) \leq M\varphi \\ (\theta_i - 1) \leq M(1-\varphi) \\ \delta \in \{0,1\} \text{ and } \varphi \in \{0,1\} \end{gathered}$$

where M is a very large positive number.

Combination of the values of the two control variables can be used to generate scenarios of target setting. For example, a value of $\delta = 0$ will imply that $z_r < 1$and that $z_r > \theta_i$. Similar scenarios can be developed for value of $\delta = 1$ and/or values of the other control variable φ. The formulated model has a mixed-integer programming format and its solution can be obtained using ordinary integer programming methods.

APPENDIX 4D

MRT & RELATIVELY EFFICIENT DMUS

The assessment of exchange rates between factors for relatively efficient units is an area of concern. Efficient units are located on the extreme points of an efficient hyperplane and thereby partial derivatives cannot be obtained for these points. Given that there is an infinite number of optimal sets of weights for efficient units, there is a difficulty in estimating productivity tradeoffs between their inputs/outputs. The presence of decision makers' preferences in M4.6 yields weights compatible to these preferences without, however, resolving the problem of multiple weights for relatively efficient units.

Using the strong complementary slackness condition (SCSC) discussed earlier, one can obtain at least one set of positive weights for all inputs/outputs. In DEA parlance this implies that from the potentially infinite sets of optimal weights for efficient units, at least one set of positive weights for all inputs/outputs can be identified. Kornbluth and Salkin (1987) introduced the notion of *incremental and decremental* shadow prices as a means of overcoming problems of multiple optimal solutions in linear programming solutions. In their method Kornbluth and Salkin recommend the solution of two LP problems where they provide the maximum and minimum values for shadow prices without affecting the current optimal solution to the problem. A more efficient process is employed here as the aim is to obtain a set of weights where all inputs/output will be weighted positively. A *max-min* LP problem is suggested therefore as formulated in M4D.1.

A Max-Min model for weights estimation (M4D.1)

$$\max_{\alpha_r,\beta_i,\gamma_r,\zeta_i} \quad \psi$$

$$\begin{array}{llll}
\sum_{i\in I_f}\beta_i x_{ij} + \sum_{i\in I_f}\zeta_i x_{ij} - \sum_{r\in O_c}\alpha_r y_{rj} - \sum_{r\in O_f}\gamma_r y_{rj} & \leq & 0 & (\forall j = 1,\ldots,n \wedge j \neq j_o) \\
\sum_{i\in I_f}\beta_i x_{ij_o} + \sum_{i\in I_f}\zeta_i x_{ij_o} - \sum_{r\in O_c}\alpha_r y_{rj_o} - \sum_{r\in O_f}\gamma_{r rj_o} & = & 0 & j_o \\
\alpha_r y_{rj_o} & \geq & P_r^+ & (\forall r \in O_c) \\
\beta_i x_{ij_o} & \geq & P_i^- & (\forall i \in I_c) \\
\gamma_r y_{rj_o} - \psi & \geq & 0 & (\forall r \in O_f) \\
\zeta_i x_{ij_o} - \psi & \geq & 0 & (\forall i \in I_c)
\end{array}$$

α_r, β_i are free variables; $\gamma_r, \delta_i \geq 0$

The formulation of M4D.1 is based on the prioritised target model in M4.6 using also a similar notation. However, some changes have been made in order to address the case of the multiple optimal sets of weights that correspond to relatively efficient units. The efficiency constraint associated with the assessed unit j_o is now represented as equality rather than inequality as j_o is an efficient unit. The role of the new variable ψ introduced in the model is to ensure that all inputs and outputs are attached a positive weight. Therefore, the optimal value of ψ is a lower positive bound for all input/output weights in the assessment.

APPENDIX 4E

GLOBAL TARGETS ASSESSMENT & PRODUCTIVITY TRADEOFFS

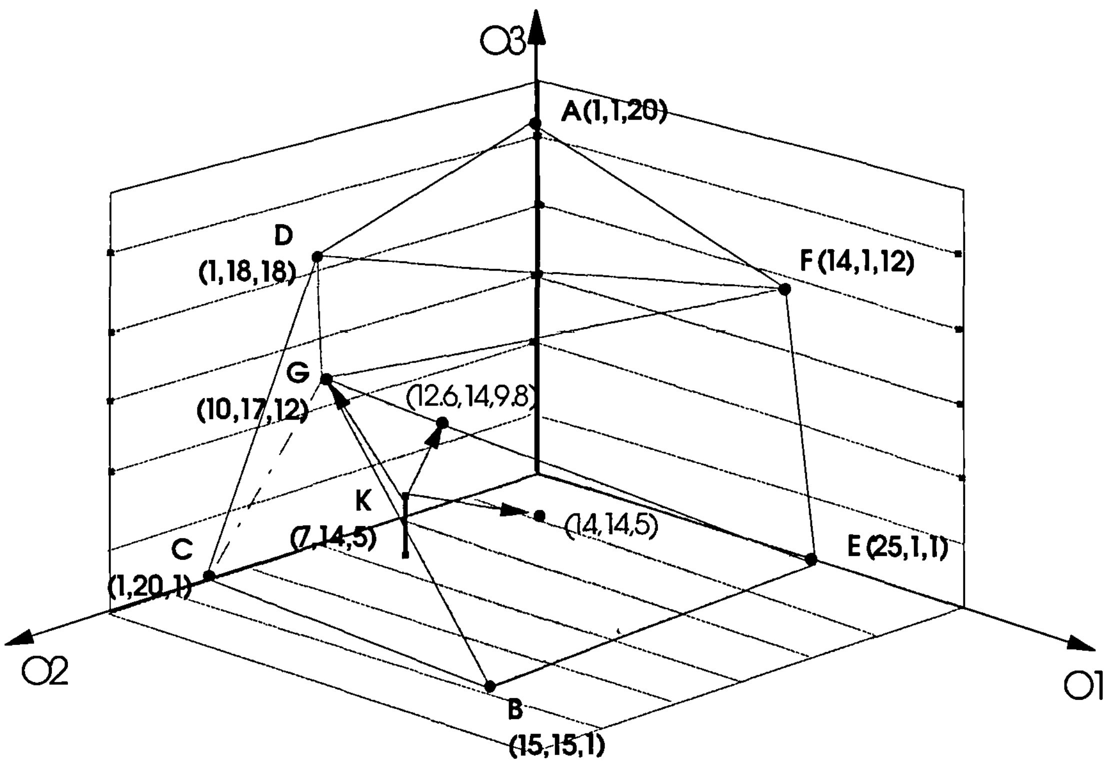
When discussing the features of the global targets models earlier on it was argued that the solution of model M4.2 yields a single set of shadow prices which represent opportunity costs at the industry level. To facilitate the discussion on the shadow prices obtained by the solution to the global targets model its dual formulation is provided in M4E.1 below.

Weights based global targets model (M4E.1)

$$\begin{aligned}
\min_{\alpha_r,\beta_i,\gamma_r,\zeta_i} \quad & \sum_{i\in I_c}\beta_i\left(\sum_{j=1}^{n}x_{ij}\right)+\sum_{i\in I_f}\zeta_i\left(\sum_{j=1}^{n}x_{ij}\right)-\sum_{r\in R_c}\alpha_r\left(\sum_{j=1}^{n}y_{rj}\right)-\sum_{r\in R_f}\gamma_r\left(\sum_{j=1}^{n}y_{rj}\right)+\sum_{r\in R_c}P_r^{+}-\sum_{i\in I_c}P_i^{-} \\
& \sum_{i\in I_c}\beta_i x_{ij}+\sum_{i\in I_f}\zeta_i x_{ij}-\sum_{r\in R_c}\alpha_r y_{rj}-\sum_{r\in R_f}\gamma_r y_{rj} \geq 0 \quad (\forall j=1,\ldots,n) \\
& \alpha_r\left(\sum_{j=1}^{n}y_{rj}\right) \geq P_r^{+} \quad (\forall r\in R_c) \\
& \beta_i\left(\sum_{j=1}^{n}x_{ij}\right) \geq P_i^{-} \quad (\forall i\in I_c) \\
& \alpha_r,\beta_i \text{ are free variables; } \gamma_r,\zeta_i \geq \varepsilon
\end{aligned}$$

The model in M4E.1 has a similar structure to that of model M4.6 but they have different objective function and "virtual" input/output constraints. In both cases, the input/output values represent the aggregate value of the corresponding input/output. This difference is effectively changing the "pricing" of weight variables in the objective function and also their corresponding constraints. The set of weights derived by the solution to model M4E.1 can be used as a basis for comparing the common set of marginal productivities with those obtained for individual DMUs when models M4.6 and M4D.1 were employed. The example of the eight DMUs listed in Table 4.2 of chapter 4) will be used again to illustrate the use of the global targets model. For illustration purposes we have reproduced the graphical representation of the three-output one input efficient frontier as was used in the main body of chapter 4 (Figure 4.3).

Figure 4.3
(reproduced from the main part of chapter 4)



Particular emphasis is given to the marginal productivities obtained from the solution of this model. Using the same methodology as before a set of seven alternative sets of targets and consequently marginal weights from the solution of M4E.1 was derived. Sensitivity analysis was employed to explore the association between alternative preference levels and the selection of efficient facets for projecting the assessed DMUs.

Table 4E.1
Global targets & exchange rates

Priority Levels for Efficiency Components (Allowable Increase-Decrease)				Productivity Tradeoffs			
Priority for output	P_1^+ (Incr., Decr.)	P_2^+ (Incr, Decr.)	P_3^+ (Incr., Decr.)	α_1/α_2	α_1/α_3	α_2/α_3	FACET
O1,O2,O3	1 (+0.79, -0.32)	1 (+3.6, -0.73)	1 (+0.54, -0.72)	1.12	1.019	0.91	G
O1	**1.79** (+2.1, -0.007)	1 +0.007, -0.5)	1 (+0.009, ∞)	2.0217	1.85	0.91	EG
O1	**3.93** (∞, -0.007)	1 (+0.009, ∞)	1 (+2.72, ∞)	4.4	1.077	0.244	EFG*
O2	1 (+0.0008, ∞)	**4.64** (+5.2, -.009)	1 (+21, -0.013)	0.242	1.02	4.21	GD
O2	1 (+0.46, ∞)	**9.91** (∞, -0.0009)	1 (+.00009 ,1)	0.166	1.45	9	GC
O3	1 (+0.0002, ∞)	1 (+14,-0.002)	**1.55** (+3.21,-.0003	1.12	0.65	0.58	GD
O3	1 (+1.89, ∞)	1 (+0.0003, ∞)	**4.77** (∞, -0.001)	3.25	0.619	0.1838	FGD*

The results listed in Table 4E.1 are organised in a way similar to Table 4.3. These results also indicate that the global target model has a similar behaviour to the one already discussed in the prioritised target model. The difference rests, however, in that the global target model will select one projection facet for all inefficient units. For example the facet EFG could be selected for projecting unit K using an appropriate set of preferences. This facet, however, was not selected when targets customised to unit K were assessed using model M4.3 (listed in Table 4.4). This indicates that the use of global target model yields industry related and not DMU related targets and therefore the estimated marginal productivities have global organisational appeal.

APPENDIX OF CHAPTER 6

The appendix of chapter 6 contains extensions concerning the development and application of the decentralised model in M6.4.

- The first case includes the mathematical models for assessing ranges of optimal weights for inputs and outputs in the solution of M6.1.
- The second, includes a modification on the way equity is represented in the decentralised model in M6.4. This concerns the definition and measurement of inequality indices as opposed to the relative need indices used in the main body of the chapter.
- The third extension, includes the formulation of a goal programming model that needs to be solved for investigating the efficiency of the solution obtained from M6.4.

APPENDIX 6A

ESTIMATING RANGES OF OPTIMAL WEIGHTS FOR INPUTS/OUTPUTS

The presence of multiple optimal sets of weights in the solution of M6.1 is very common for relatively efficient units whilst more unusual for inefficient units. Therefore, the selection of sets of weights to represent relatively efficient units in the solution of the DTP model can be highly judgmental. The alternative formulations in M6.5 and M6.5a require specification of the range of optimal weights for each input and output in the solution of M6.1. These weights can be obtained following the two-phase approach discussed next.

Let assume that optimal sets of weights were obtained by the solution of M6.1 for unit j_O. The formulation in M6A.1 can be used as the basis to generate a sequence of linear programmes that will be used to obtain upper and lower bounds for weights on inputs/outputs. The model in M6A.1 is demonstrated for output r of DMU j_O.

$$\alpha^m_{rj_o} = \{\min \quad \alpha_r | M6A.1\} \text{ and } \alpha^M_{rj_o} = \{\max \quad \alpha_r | M6A.1\}$$

$$\begin{array}{llll} s.t & & & \\ \sum_{i \in I_c} \beta_i x_{ij} + \sum_{i \in I_f} \zeta_i x_{ij} - \sum_{r \in O_c} \alpha_r y_{rj} - \sum_{r \in O_f} \gamma_r y_{rj} & \geq & 0 & (\forall j \neq j_o) \\ \sum_{i \in I_c} \beta_i x_{ij_o} + \sum_{i \in I_f} \zeta_i x_{ij_o} - \sum_{r \in O_c} \alpha_r y_{rj_o} - \sum_{r \in O_f} \gamma_r y_{rj_o} & = & t^*_{j_o} & \\ \alpha_r y_{rj_o} & \geq & P^+_r & (\forall r \in O_c) \\ \beta_i x_{ij_o} & \geq & P^-_i & (\forall i \in I_c) \\ \alpha_r, \beta_i \text{ are free variables; } \gamma_r, \zeta_i \geq \varepsilon. & & & \end{array} \qquad \text{(M6A.1)}$$

where $t^*_{j_o}$ is the optimal solution obtained from M6.1 for DMU j_O.

The current formulation of M6A.1 requires solution to two linear programming problems that differ on the direction of the objective function. The first problem yields an optimum set of weights where output r takes the minimum weight whilst the second problem yields an optimum set of weights where output r takes the maximum weight. Similar pairs of linear problems can be solved to derive the range of weights for all inputs ($\beta^m_{ij}, \beta^M_{ij}$) and outputs ($\alpha^m_{rj}, \alpha^M_{rj}$) of all DMUs with multiple optimal solutions.

APPENDIX 6B

DECENTRALISED PLANNING OBJECTIVES & INEQUALITY

Hitherto, the resource allocation model in M6.4 used relative need criteria, expressed in M6.4d, in order to incorporate equity as a resource allocation objective. An additive social welfare value function was employed to obtain an index of relative need of individual DMUs. This index is then applied individually to the allocation of each global controllable input Gx_i. According to this specification, equity is achieved if all DMUs are allocated resources in proportion to their relative need score. This objective is pursued in the resource allocation process taking into account the tradeoffs between the satisfaction of the other two objectives, namely effectiveness and efficiency.

There is a considerable theoretical and empirical literature that provides more rigorous definition of equity as a planning objective. Savas (1978) argues for a classification of equitable allocation methods into four different types, namely equal payments, equal outputs, equal inputs, equal satisfaction of demand. The economic literature also has substantial contribution to the equity debate through the notion of *inequality* see Allison (1978), Atkinson (1970), Bartels and Nijkamp (1976).

Inequality in welfare economics is typically defined using alternative relative distance functions, see Gini coefficient, Theil inequality function (1967), and simple statistical coefficient of variation of the allocation of resources to different activity centres. In a comparative study employed by Allison (1978) it was argued that the Gini coefficient is the most commonly used index of inequality without, however, offering any substantial benefits over the other two indices. In the context of this study we do not pursue any further the debate on the choice of inequality indices. The analysis will concentrate, therefore, on how one index, namely the Gini coefficient, can be used as an equity criterion in the target-based planning model in M6.4.

Let us assume that there is a distribution of public services to a set of n operating units (activity centres). We shall denote S_j the set of service units (resources) received by unit j. It is assumed for the time being that the service units are expressed in the same values (e.g. pounds). We shall also denote Q_j the number of "equity units" contained in unit j. Equity units represent factors (e.g. population) that will be used as criteria for the allocation of services. The variables used to represent the service S_j and equity Q_j units will typically have a multidimensional nature and their representation in the Gini coefficients need to be discussed next.

Berne and Stiefel (1984) provide a mathematically tractable expression of the Gini coefficient that is presented in M6B.1.

$$G = \frac{\sum_j \sum_{k>j} \left| q_k S_j - q_j S_k \right|}{\sum_j S_j} \tag{M6B.1}$$

where

S_j is the number of service units received by unit j,

Q_j number of "equity units" of the j^{th} unit,

$q_j = Q_j \big/ \sum_j Q_j$ is the proportion of "equity units" in the area of unit j,

$\sum_j Q_j$ is the total number of "equity units" for all j units.

The mathematical representation of M6B.1 is discussed next in order to give a managerial interpretation to its formulation. The formula in M6B.1 represents the aggregate pairwise comparisons of combinations of equity and allocated resources of activity centres. In the numerator we have all differences between the resources (S_j) allocated to, say, DMU j weighted by the relative need (q_k) of, say, DMU k and the resources (S_k) allocated to DMU k weighted by the relative need (q_j) of DMU j. The denominator contains the total number of allocated resources which serves as a standardisation coefficient. The inequality index G, therefore, yields a numerical estimate of the extent to which the allocation of service units to activity centres is in proportion to their relative need.

One of the mathematical properties of the Gini coefficient in M6B.1 is that it reflects absolute inequality aversion rather than relative inequality aversion. In other words, increasing the level of service per equity unit by a constant amount decreases the Gini coefficient which indicates greater equity. A multiplication of the level of service per equity unit by a constant will have no effect on the Gini coefficient.

The Gini coefficient defined in M6B.1 can be applied directly to cases with service and equity units being represented by single dimension variables. For example, this could be applied to a case where the budget constitutes the service units and population constitutes the equity units respectively. In cases, however, where the allocated resources reflect more composite factors e.g. personnel, equipment, recurrent grants, etc. and the equity units reflect a variety of relative need criteria the formulation in M6B.1 need to be extended.

The formulation in M6B.1 will be extended next to assess inequality among DMUs with multiple inputs-outputs. The extension succeeds to aggregate service units (multiple inputs) into a single dimension without any presumptions on the prices of the multiple inputs of each DMU. This task is pursued making use of the earlier methodology of the network representation of MUOs as was proposed in the introduction of Chapter 6.

The "service units" can thereby be aggregated using the notion of "input flow", defined earlier in Chapter 6, by means of the weighted sum, $\sum_{i\in I_c} \beta_i^* \phi_{ij}$, of controllable inputs of individual activity centres (DMUs) in model M6.2. The assessment of "equity units", however, will remain as the weighted proportion of relative need of individual DMUs defined already in M6.4d. A modified formula of inequality measurement is provided in M6B.1a that seeks to incorporate the two enhancements made above.

$$G = \frac{\sum_j \sum_{k>j} \left| q_k \sum_{i\in I_c} \beta_{ij}^* \phi_{ij} - q_j \sum_{i\in I_c} \beta_{ik}^* \phi_{ik} \right|}{\sum_{i\in I_c} \sum_j \beta_{ij}^* \phi_{ij}} \qquad \text{(M6B.1a)}$$

and

$$q_j = \sum_{e\in E} w_e \frac{x_{ej}}{\sum_{j=1}^{n} x_{ej}} \qquad \text{(M6B.1b)}$$

Where,

q_j, q_k are composite equity units for the j[th] and k[th] DMU correspondingly estimated by the weighted multiattribute value function in M6B.1b,

β_{ij}^* is the optimal weight factor estimated by the prioritised target model (M6.1),

x_{ej} is the <u>observed</u> value of the e[th] equity criterion of DMU j,

ϕ_{ij} is the variable associated with the i[th] input quantity of DMU j,

w_e is the weight of importance of the e[th] equity criterion.

The formulation of the Gini coefficient in M6B.1a has the same managerial interpretation as in M6B.1. The only part that needs to be discussed concerns the estimation of the various multivariate components of the formulae. The representation of the input flow (previously called service units) is made using the network flow balance equations discussed in the development of the model in M6.4.

It is noteworthy, however, that the use of the network flow components for individual DMUs is valid so long as they are estimated via the prioritised DEA model discussed in detail in chapter 4 and reproduced in chapter 6 (M6.1). Possible use of the radial CRS/VRS models (see M2.8 in chapter 2) for obtaining these input flow components will give non-comparable flow equations for individual DMUs due to the input or output standardisation[2] that takes place prior to solving these DEA problems.

Another issue concerning the multivariate Gini coefficient defined in M6B.1 emanates from the estimation of the technological coefficients for controllable inputs. These coefficients are obtained from the prioritised target model in M6.1 which may yield multiple optimal sets of weights for efficient DMUs. The presence of multiple optimal sets of weights does not affect the aggregate value of the inflow equation $\sum_{i \in I_c} \beta^*_{ij} \phi_{ij}$, although it will affect the contribution of individual inputs $\beta^*_{ij} \phi_{ij}$ to that value. The case of the multiple optimal sets of weights, although it does not affect M6.B1, can be addressed using the range of optimal weights discussed earlier in Appendix 6A.

The target-based planning model revised

Having extended the Gini coefficient in a multi-input multi-output production space (M6A.1a) the next step seeks to incorporate this formula in DTP model in M6.4. Inequality is usually perceived as an undesirable effect in an allocation/distribution system and, therefore, inequality as a resource allocation objective needs to be minimised. The formulation of equity introduced in M6B.1 cannot be used to define specific target values of inequality that would enable its representation in a goal programming sense[3].

The representation of equity in the DTP model of M6.4 leads to its modified version in M6B.2.

---

[2] The typical DEA model is solved by standardising the weighted sum of inputs or outputs of assessed DMUs to be equal to 1. The fact that a different standardisation takes place for each DMU makes the comparisons between, say, the input flow component $\sum_i a_{ij} x_{ij}$ of alternative DMUs meaningless.

[3] The latter was possible when equity was defined along the lines of relative need in M6.1d.

Decentralised target-based planning & inequality (M6B.2)

$$\underset{\psi_{rj},\phi_{ij},t_j,D_i,D_r}{Min}\left\{\sum_{r\in O_c}\frac{P_r^+D_r^+ + P_r^-D_r^-}{Gy_r}+\sum_{i\in I_c}\frac{P_i^+D_i^+ + P_i^-D_i^-}{Gx_i},\ \sum_{j=1}^{n}P_jt_j\right\} \qquad \text{(M6B.2a)}$$

$$\underset{\phi_{ij},\psi_{rj}}{Min}\quad G=\frac{\sum_j\sum_{k>j}\left|q_k\sum_{i\in I_c}\beta_{ij}^*\phi_{ij}-q_j\sum_{i\in I_c}\beta_{ik}^*\phi_{ik}\right|}{\sum_{i\in I_c}\sum_j\beta_{ij}^*\phi_{ij}} \qquad \text{(M6B.2b)}$$

Subject to

Effectiveness	$\sum_{j=1}^{n}\psi_{rj}+D_r^- - D_r^+ = Gy_r$	$(\forall r\in O_c)$	**(M6B.2c)**
	$\sum_{j=1}^{n}\phi_{ij}+D_i^- - D_i^+ = Gx_i$	$(\forall i\in I_c)$	
Efficiency	$\sum_{i\in I_c}\phi_{ij}\beta_{ij}^* - \sum_{r\in O_c}\psi_{rj}\alpha_{rj}^* - t_j = F_j^*$	$(\forall j)$	**(M6B.2d)**
	$F_j^* = -\sum_{i\in I_f}x_{ij}\zeta_{ij}^* + \sum_{r\in O_f}y_{rj}\gamma_{rj}^*$		
Policy constraints	$L\phi_{ij}\le\phi_{ij}\le U\phi_{ij}$	$(\forall i\in I_c)$	**(M6B.2e)**
	$L\psi_{rj}\le\psi_{rj}\le U\psi_{rj}$	$(\forall r\in O_c)$	
Inequality component	$q_j=\sum_{e\in E}w_e\frac{x_{ej}}{\sum_{j=1}^{n}x_{ej}}$	$(\forall j)$	**(M6B.2f)**
	$\phi_{ij},\psi_{rj},D_i^{+,-},D_r^{+,-},t_j\ge 0$		

The notation of model M6B.2 is as defined in M6.4 and M6B.1 above. The nature, however, of M6B.2 is a multi-objective one (bi-criteria). The first objective constitutes the part of the goal programming model formulated in M6.4 that corresponds to the satisfaction of the objectives of effectiveness and efficiency. The second objective represents the satisfaction of the objective of equity by minimising the aggregate inequality in the allocation of resources. The constraints of the problem (M6B.2c-M6B.2f) correspond to the representation of effectiveness, efficiency, policy constraints and the inequality component.

A similar problem faced in the formulation engineered by Mandell (1991) was solved using *frontier-generating* techniques from the multi-objective programming literature. In brief, this solution process anticipates that, in bi-criteria problems, one can generate the frontier of non-dominated solutions simply by converting one of the objective functions to an inequality and then generate optimal values for the other objective by parametric variation of the value of the introduced constraint (ε-method Cohon (1978)). Having generated the non-dominated frontier for two objectives a graphical interface can be employed for selecting the solution with the most preferred tradeoffs. The adoption of a solution approach for the multi-objective problem in M6B.2 is left to be context dependent as there are other methodologies one can adopt to derive solutions to the problem.

APPENDIX 6C

TESTING FOR NON-INFERIOR SOLUTIONS IN THE DTP MODEL IN M6.4

As Min and Storbeck (1991) emphasise, there is no unanimous agreement concerning the conditions for non-inferior (see definition in the main body of chapter 6) solutions in goal programming. It is widely recognised, however, that one needs to be aware of the possibility of obtaining inferior solutions from certain forms of goal programming models. It is worth mentioning that a simple inspection on the very wide goal programming literature (see Zanakis and Gupta (1985) and Romero (1991)) shows that the possibility of inferior solutions in goal programming tends to be neglected by the vast majority of published research.

A number of remedies have been proposed providing alternative methods for obtaining efficient solutions from goal programming models. The application of these methods on the solution process of the decentralised planning model is illustrated using the model suggested by Hannan (1981) and also applied by Thanassoulis and Dyson (1992) in a DEA context.

Inferior solutions in a goal programming problem are possible due to the simultaneous presence of under/over achievement deviation variables for some of the goals of the problem, see Zeleny (1982), Romero (1991). The formulation of M6.4 contains two direction deviation variables in the set of goal constraints used to reinforce the equity criterion. Let assume that a solution $\left(D_i^{-*}, D_r^{+*}, \varepsilon_{ij}^{-*}, \varepsilon_{ij}^{+*}, t_j^*\right)$ was found to the M6.4 problem.

The linear programming model in M6C.1 can be solved to investigate the presence of non-efficient solutions in the solution of the DTP model.

Testing for non-inferior solutions in the DTP (M6C.1)

$$\underset{x_{ij},y_{rj},c_{icj},L_i,L_r}{Min} \quad \sum_{r\in O_c} L_r + \sum_{i\in I_c} L_i + \sum_j \sum_{i\in I_c} c_{ij}$$

Subject to

$$\sum_{j=1}^{n} \psi_{rj} + L_r = Gy_r + D_r^{+*} - D_r^{-*} \qquad (\forall r \in O_c)$$

$$\sum_{j=1}^{n} \phi_{ij} - L_i = Gx_i + D_i^{+*} - D_i^{-*} \qquad (\forall i \in I_c)$$

$$\sum_{i\in I_c} \beta_i^* \phi_{ij} - \sum_{r\in O_c} \alpha_r^* \psi_{rj} = F_j^* + t_j^* \qquad (\forall j)$$

$$F_j^* = \sum_{i\in I_f} \zeta_i^* x_{ij} - \sum_{r\in O_f} \gamma_r^* y_{rj}$$

$$\phi_{ij} + c_{ij} = Gx_i * \left(\sum_{e\in E} w_e \frac{x_{ej}}{\sum_{j=1}^{n} x_{ej}} \right) + \varepsilon_{ij}^{+*} - \varepsilon_{ij}^{+*} \qquad (\forall i \in I_c)$$

$$L\phi_{ij} \le \phi_{ij} \le U\phi_{ij} \qquad (\forall i \in I_c)$$

$$L\psi_{rj} \le \psi_{rj} \le U\psi_{rj} \qquad (\forall r \in O_c)$$

$$\phi_{ij}, \psi_{rj}, L_r, L_i, c_{ij} \ge 0$$

The notation of M6C.1 is similar to models M6.4 with L_r, L_i, c_{ij} being a new set goal deviation variables.

The direction chosen for the extra goal variables in the model emanates from the implicit assumption concerning the aim of output goals' overachievement and input goals' underachievement. If the optimal solution of M6C.1 yields $L_r = L_i = c_{ij} = 0$ that would imply that no further improvements are feasible to the solution obtained by M6.4. Otherwise, the solution obtained by M6.4 was inefficient and adjustments in line with the solution obtained by M6C.1 need to be made to render the solution efficient.

APPENDICES OF CHAPTER 7

The appendices related to chapter 7 are mainly concerned with the provision of background information of the Brewing industry in the UK and the descriptive statistics of the data set used in our case study. Some statistical results concerned with the assessment of pubs' performance are also discussed.

APPENDIX 7A

The Brewing industry in the UK

Over the period 1987-1993 the brewing industry in the UK has been subject to intense pressure for reorganisation. The adverse general economic climate and the governmental attempt to reinforce competition within the brewing industry are the main forces for reorganisation.

The most important characteristic of the brewing industry in the UK is the decline in its market size. Table 7A.1 shows the marginal decline in the number of public houses in the UK in 1990 and 1991 following a number of years of modest growth. Turnover of public houses also decreased in 1991. This decline in turnover can be attributed to the general decline in disposable income, the 1989-1992 recession, to price rises in the industry above inflation and to the increasingly competitive role played by off-trade[4].

Table 7A.1

Number and Turnover of Public houses in the UK, (1987-1991)

Year	No. of outlets	Turnover (£ Million inc. VAT)	Alcohol sales as (%) of total sales
1987	68500	8274	73
1988	68825	8716	71
1989	69450	9712	70
1990	69300	10295	67
1991	69000	9980	63

Source: Brewers' Society and Business Statistics data

4 This relates to trading stores that hold licences for selling alcohol for at home consumption.

The decline in the market's size, as well as the increasing costs for maintaining and improving the state of repair of public houses are the main reasons for price rises in pubs (e.g. 11% tax free increase in 1991). The price increases for compensating lost revenue from reduced volume of sales has very limited appeal in the long run. Increased prices result in further customers' (sales) losses which are switched to the off-trade market. The decline of the alcohol sales to the aggregate turnover composition from 73 to 63 percent indicates the growing interest of public houses in providing full catering facilities.

The brewing market in the UK is an oligopoly. Over half of the total number of public houses in the UK (see Table 7A.1) are controlled by less than ten major brewers whilst the top four brewers have over 6000 pubs each, accounting for 46 % of pubs in the UK.

There are three main types of pubs in the UK based on their type of management; tenancies, managed pubs and free pubs. Under the tenancies or tied system tenants rent their pubs from the brewery, and they are not free to buy from a supplier of their choice. Managed pubs are under more direct control of the brewery as the managers are brewery employees. Free houses are run by independent management that are, in theory, free from ties to individual breweries. This independence is, however, limited due to various financial links (e.g. loans, supply agreements) with the major brewers.

Despite the downturn in alcoholic drink consumption (see Table 7.1), alcoholic products still account for over 60% of the turnover of the average public house. However, the importance of alcoholic drink sales to public houses is declining as shown in Table 7.A1. Evidently, the decline in the consumption of alcohol over time created a favourable era towards broadening traditional services provided by public houses. Provision of food and catering services account for over 30% of the turnover of the average public house.

Public houses' emphasis on food services, however, increase the pressure on brewers to invest a higher proportion of capital in their retail activities rather than in their manufacturing activities. The proportion of investments devoted to manufacturing activities has fallen from 19% in 1986 to 15% in 1990 whilst the proportion of investments devoted to retailing activities has risen from 62% in 1986 to 72% in 1990 (by value).

Another important factor related with the management of public houses is concerned with the variety of types of public houses depending on their location, type of services they provide, targeted clientele, etc. There is no unique classification system of public houses in the UK and each brewery uses its own list of pub categories.

The pubs used in this study constitute the so-called "broad based" pubs. These pubs rely on local and passing customers whilst their services vary from drink to food oriented.

Broad based pubs are on a transition period as central management sought to increase the range and quality of catering facilities in accordance with the general trend discussed earlier.

Descriptive and statistical analysis of the input-output set

The location characteristics of pubs in the study vary considerably. The location effects on pubs' performance are examined later in the chapter. Table 7A.2 classifies the pubs based on three location characteristics.

Table 7A.2

Location characteristics (average) of pubs within 2.5 miles radius

Location	No. of Pubs	Turnover (£)	Parking places (No.)	State of repair [1,15]	Potential Custom	Alcohol Consumption (barrels of beer)	Compe -titors	Bar Area (ft^2)
Industrial Estate	65	166640	28	11	9718	104902	10	1241
Main road	46	182461	20	10	11800	122501	22	1323
Suburban Areas	33	182064	17	9	13627	150218	11.4	1351
Total	154	171936	23	10.5	10523	112794	11	1255

Table 7.2 shows a summary of the average characteristics of pubs located into three main areas. The data provided refer to proposed determinants of market efficiency discussed earlier on. Pubs located in *industrial estates* seem to have smaller turnover than the rest of pubs as a result of the smaller number of potential customers, as well as lower consumption of alcohol in their surrounding area. Competition is variable with *main-road* pubs facing the most intense competition as they have on average 22 pubs within their trade area. Surprisingly enough, *industrial estate* pubs have the highest state of repair as a result of perhaps higher investment by the brewery.

A linear regression model was used to examine the statistical association between the input factors used in Figure 7.6 and the generated turnover. These results are provided in Table 7A.3 below.

Table 7A.3
Turnover as a function of input variables

Regression variable	Turnover (£)
Bar area (ft^2)	34.64*
Number of car park facilities	647.00*
State of repair of pubs	4928.00*
Consumption of alcohol (barrels) per premise of alcohol trade in 2.5 miles radius	0.0067*
Number of potential customers in 2.5 miles radius	3.95*
Average households' income in 2.5 miles radius	7.56*
Dummy variable (1, 0) for pubs with and without food sales	42147.30*
Constant term	- 103378.00*
R^2	55 %

(*) Significant at the 5% level

The regression analysis model was employed to "validate" the statistical significance of the factors proposed as determinants of market efficiency of pubs. The explained variation of the total turnover was 55%. All input factors employed in the regression model were found statistically significant at a 5% level. The explanatory factors in the regression model do not coincide with those exhibited in the influence diagram in Figure 7.6 (in the main body of chapter 7). Some variables were modified (e.g. income in the surrounding area) whilst some others were merged into one variable (e.g. consumption of alcohol per licensed premise) in order to avoid spurious results in the regression model due to multicollinearity.

The internal factors of market efficiency have on average a positive association with the turnover of pubs. For instance, each square foot of bar area has an average return of £ 34.64 whilst each extra parking space contributes on average £647. State of repair has an average contribution of £4928 for each extra point in the qualitative scale of measurement [1, 15]. One could estimate, perhaps, the per unit scale cost of improving the state of repair of pubs and contrast that with the average returns.

External factors have also positive influence on the turnover of pubs. For instance each potential customer of pubs has on average a contribution of £3.95 to the pubs' turnover. Evidence from Table 7A.3 also indicates that on average public houses with food services have £42,147 extra turnover.

A better fit of the model could be explored by employing different functional forms linking turnover with market efficiency factors. The latter, however, would open a debate on the identification of the most appropriate functional form (e.g. Cobb-Douglas, Translog, Quadratic, etc.) to fit the data which was not one of the intentions of the current study. Regression analysis is employed here to shine some light on the statistical association between the various factors of market efficiency and the generated turnover of pubs. Issues related with the use of regression analysis as a performance measurement tool are discussed in more detail in Thanassoulis (1993) and Banker *et al.* (1986).

Notwithstanding, the model in Table 7A.3 shows that the suggested determinants of market efficiency are at least statistically significant. Statistical significance, however, is not the only criterion for identifying factors that affect the performance of operating units. Varian (1984) for instance makes use of the term "economic significance" to emphasise the presence of non-statistically significant factors which are often sacrificed by the sole use of statistical yardsticks in estimating production functions. To elaborate further this remark, one should realise that if a factor, say, consumption of alcohol in the surrounding area was not statistically significant in a regression model this could imply a non-linear association and/or an excess underutilisation of the market potential by the pubs in the sample. The specification of causal models for assessing market efficiency, therefore, should combine managerial judgement and statistical analysis.

APPENDIX 7B

The association between pubs' size and efficiency

The association between pubs' size and performance was investigated using the chi-square test described in Table 7B.1. Pubs were classified in this Table using the quartile scores for pub size and aggregate market efficiency.

Table 7B.1

Distribution of pubs by size & aggregate market efficiency

	Pubs with bar area (ft^2)					
Pubs with efficiency (%)	≤ Qr.1= 900		Qr.1- Qr.3		≥ Qr.3 = 1590	
	Obs.	*Exp.*	Obs.	Exp.	Obs.	Exp.
0 - 60	4	*11*	21	*17*	13	*9*
61 - 94	25	*20*	28	*33*	20	*18*
95 - 100	15	*13*	22	*11*	6	*12*
X^2= 11.14, significant at 5% level with 4 d.f.						

The statistical test in Table 7B.1 shows that there is a statistically significant association between size and aggregate market efficiency of assessed pubs. Indeed, it was found that pubs with smaller bar area tend to outperform relatively larger pubs. This can lead to unrealistic performance comparisons as there is no evidence indicating that all being equal if we double the size of a pub its turnover should be doubled. For example, the large size of pubs can, sometimes, be a counter attractive factor for certain types of customers that prefer smaller and quieter places for entertainment.

APPENDIX OF CHAPTER 8

INVESTIGATING THE RATIONALE OF THE IMPROVED SCALE SIZE (ISS) MODEL

In this part we provide some further information regarding the rationale of the Improved Scale Size model that was developed in the main body of chapter 8 for developing planning scenarios in for-profit outlets. The main emphasis here is to describe the effects of M8.3 on the assessed efficiency and hence efficient targets of inefficient units. To facilitate this discussion we use the dual formulation (defensive model) of M8.3 which is provided in M8A.1.

Improved Scale Size targets defensive model **M8A.1**

$$\underset{\alpha_r,\beta_i,\gamma_r,\zeta_i}{Min} \quad \sum_{i\in I_c} \beta_i x_{ij_o} \qquad \text{M8A.1a}$$

$$\sum_{r\in O_c} \alpha_r y_{rj_o} = 100 \qquad \text{M8A.1b}$$

$$\sum_{i\in I_c} \beta_i x_{ij} - \sum_{r\in O_c} \alpha_r y_{rj} + \sum_{i\in I_f} \zeta_i \left(x_{ij} - x_{ij_o}\right) - \sum_{r\in O_f} \gamma_r \left(y_{rj} - y_{rj_o}\right) \geq 0 \quad \forall j \qquad \text{M8A.1c}$$

$$\alpha_r, \beta_i, \gamma_r, \zeta_i \geq \in$$

Where,

x_{ij}, y_{rj} are quantities of input i and output r of DMU j,

I_c, O_c are subsets of controllable inputs and outputs respectively,

I_f, O_f are subsets of uncontrollable inputs and outputs respectively.

The assessment of DMU j_O includes in the objective function (M8A.1a) and the standardisation constraint (M8A.1b) only the controllable inputs and outputs (I_c , O_c). Furthermore, the constraints that link the weighted sum of inputs to the weighted sum of outputs for each DMU j differ from the constraints used in a typical DEA assessment (see M2.8 in chapter 2). These differences are discussed in more detail next.

The uncontrollable input and output levels of each DMU are adjusted by subtracting the observed levels of the assessed unit j_O (This is why the objective function and the standardisation constraint do not include these inputs at all). This standardisation affects the very mechanism of the DEA comparisons as follows. The assessed unit j_O will be rated relatively inefficient if at least one of the remaining DMUs has its constraint in M8A.1c binding in the optimal solution of j_O. The adjustments of uncontrollable inputs/outputs of each DMU j can affect their likelihood to have their constraints binding in a positive or negative way. Let us consider the case of unit j and input i:

1. For unit j, if $x_{ij} < x_{ij_o}$ then the difference is $\tilde{x}_{ij} = x_{ij} - x_{ij_o} < 0$ and thus the revised input $\tilde{x}_{ij}$ of unit j is taken as an output during the assessment of unit j_O. The latter implies that it is more likely for unit j to balance its inputs with its outputs and therefore to have its constraint binding and thus j_O inefficient. Thus, units with smaller uncontrollable inputs are compensated by the model for the difference between their values and the values of the assessed unit j_O.

2. On the other hand if $x_{ij} > x_{ij_o}$ then the difference $\tilde{x}_{ij} = x_{ij} - x_{ij_o} > 0$ and therefore the revised input $\tilde{x}_{ij}$ is taken as an inputs during the assessment of unit j_O. The latter implies that for units with higher uncontrollable inputs to those of the assessed unit j_O their difference operates as an input which needs to be considered (weighted) in order to have their constraints binding.

Similar arguments can be made for the effect of the uncontrollable outputs on the assessment of the efficiency of unit j_O. Undoubtedly, the situation will become more complicated in cases with simultaneous presence of uncontrollable inputs and outputs where some of them have higher and some smaller input or output values than the assessed unit j_O.